

eHEALTH BEYOND THE HORIZON – GET IT THERE

Studies in Health Technology and Informatics

This book series was started in 1990 to promote research conducted under the auspices of the EC programmes' Advanced Informatics in Medicine (AIM) and Biomedical and Health Research (BHR) bioengineering branch. A driving aspect of international health informatics is that telecommunication technology, rehabilitative technology, intelligent home technology and many other components are moving together and form one integrated world of information and communication media. The complete series has been accepted in Medline. Volumes from 2005 onwards are available online.

Series Editors:

Dr. J.P. Christensen, Prof. G. de Moor, Prof. A. Famili, Prof. A. Hasman, Prof. L. Hunter, Dr. I. Iakovidis, Dr. Z. Kolitsi, Mr. O. Le Dour, Dr. A. Lymberis, Prof. P.F. Niederer, Prof. A. Pedotti, Prof. O. Rienhoff, Prof. F.H. Roger France, Dr. N. Rossing, Prof. N. Saranummi, Dr. E.R. Siegel, Dr. P. Wilson, Prof. E.J.S. Hovenga, Prof. M.A. Musen and Prof. J. Mantas

Volume 136

Recently published in this series

- Vol. 135. T.B. Grivas (Ed.), The Conservative Scoliosis Treatment – 1st SOSORT Instructional Course Lectures Book
- Vol. 134. B. Blobel, P. Pharow and M. Nerlich (Eds.), eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge – Global Experts Summit Textbook
- Vol. 133. J. Hammer, M. Nerlich and S. Dendorfer (Eds.), Medicine Meets Engineering – Proceedings of the 2nd Conference on Applied Biomechanics Regensburg
- Vol. 132. J.D. Westwood, R.S. Haluck, H.M. Hoffman, G.T. Mogel, R. Phillips, R.A. Robb and K.G. Vosburgh (Eds.), Medicine Meets Virtual Reality 16 – parallel, combinatorial, convergent: NextMed by Design
- Vol. 131. R. Latifi (Ed.), Current Principles and Practices of Telemedicine and e-Health
- Vol. 130. J.I. Westbrook, E.W. Coiera, J.L. Callen and J. Aarts (Eds.), Information Technology in Health Care 2007 – Proceedings of the 3rd International Conference on Information Technology in Health Care: Socio-technical Approaches
- Vol. 129. K.A. Kuhn, J.R. Warren and T.-Y. Leong (Eds.), MEDINFO 2007 – Proceedings of the 12th World Congress on Health (Medical) Informatics – Building Sustainable Health Systems
- Vol. 128. P.J. Murray, H.-A. Park, W.S. Erdley and J. Kim (Eds.), Nursing Informatics 2020: Towards Defining Our Own Future – Proceedings of NI2006 Post Congress Conference
- Vol. 127. L. Bos and B. Blobel (Eds.), Medical and Care Compunctics 4

eHealth Beyond the Horizon – Get IT There

Proceedings of MIE2008

The XX1st International Congress of the European Federation for Medical Informatics

Edited by

Stig Kjær Andersen

*Department of Health Science and Technology, Aalborg University,
Aalborg, Denmark*

Gunnar O. Klein

*Department of Microbiology, Tumour and Cell Biology, Karolinska Institute,
Stockholm, Sweden*

Stefan Schulz

*Department of Medical Informatics, Freiburg University Medical Centre,
Freiburg, Germany*

Jos Aarts

*Institute of Health Policy and Management, Erasmus University,
Rotterdam, the Netherlands*

and

M. Cristina Mazzoleni

Scientific Direction, Salvatore Maugeri Foundation IRCCS, Pavia, Italy

IOS
Press

Amsterdam • Berlin • Oxford • Tokyo • Washington, DC

© 2008 Organizing Committee of MIE 2008.

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without prior written permission from the publisher.

ISBN 978-1-58603-864-9

Library of Congress Control Number: 2008926215

Publisher

IOS Press

Nieuwe Hemweg 6B

1013 BG Amsterdam

Netherlands

fax: +31 20 687 0019

e-mail: order@iospress.nl

Distributor in the UK and Ireland

Gazelle Books Services Ltd.

White Cross Mills

Hightown

Lancaster LA1 4XS

United Kingdom

fax: +44 1524 63232

e-mail: sales@gazellebooks.co.uk

Distributor in the USA and Canada

IOS Press, Inc.

4502 Rachael Manor Drive

Fairfax, VA 22032

USA

fax: +1 703 323 3668

e-mail: iosbooks@iospress.com

LEGAL NOTICE

The publisher is not responsible for the use which might be made of the following information.

PRINTED IN THE NETHERLANDS

Preface

The first part of the MIE 2008 conference theme – *eHealth Beyond the Horizon* – highlights the expectations for the future of eHealth and raises the question: What sort of developments in eHealth services can we imagine emerging above the horizon in the years to come? We have received a good number of high quality papers giving different perspectives of this future, some of them already available today in pilot scale, some of them outlined in visions.

The second part of the theme – *Get IT there* – has triggered a large number of papers describing how to create, evaluate, adjust and deliver products and deploy services in health care organizations for the necessary information technology as a basis for the eHealth applications that are essential in order to respond to the challenges of the health systems.

We, as society, are facing great challenges for our health systems. Our national health systems will have to treat proportionately more people due to the change in age distribution, with more illnesses, higher expectations, and more expensive technologies based on the biomedical scientific progress, but using proportionally less money and fewer workers. In addition, the increased mobility between countries within the European Union and a large number of immigrants from the war zones of the world creates challenges. On the other hand, the increased global co-operation and willingness of the countries to co-operate and address the possibilities for cross-border care – e.g. for highly specialised care and centres of excellence – give hope. The conference will address the challenges we face, in order to cope with these changes. We cannot know for sure what is below the horizon, but we know that enormous possibilities lie in eHealth.

As researchers, we will share our results and critical evaluation of new methods and possibilities. We encourage an open and constructive scientific discussion that is global in nature. We are very happy that this year MIE has also attracted a good number of non-European participants from all continents.

We, as stakeholders, want the medical informatics society to contribute substantially to the agenda and make the roadmap for the future health services. For several decades information technology has been applied in healthcare in interplay among stakeholders, organizations, and applications. Let us identify what makes a difference and brings the best of the past into the future.

When we grasp the future and have to fulfil the expectations, we have to ensure that the health applications are reliable, interoperable, sustainable, efficient, effective, and ubiquitous. The vital questions we therefore face are: What form of health systems, i.e. can we envisage which are sustainable in the long run, and how do we set about building them? How do we engage in optimizing the health services, given the complexity of current organizations, current skills and resource shortages, and the fundamental barriers that exist everywhere to undertake radical change?

Thoughts, visions, and objectives on the European arena have been brought forward by the European Commission with an action plan for a European eHealth Area: eHealth – making healthcare better for European citizens acknowledging that eHealth and interoperability can offer a number of golden opportunities. The 7th Framework Program will contribute to the realization of these visions.

This conference celebrates that 30 years have gone by since the first conference was organised by the European Federation of Medical Informatics – EFMI. The conference this year is organised by the Swedish Federation of Medical Informatics but the scientific evaluation and organisation of the conference has been conducted by a broad European Scientific Programme Committee in co-operation with the EFMI council. The following persons were members of the SPC: Stig Kjær Andersen (Chair), Jos Aarts, Marie-Christine Jaulent, Gunnar Hartvigsen, Jacob Hofdijk, Dipak Kalra, Gunnar Klein, Sabine Koch, Yunkap Kwankam, Victor Maojo, M. Cristina Mazzoleni, George Mihalas, Göran Petersson, Martin Rydmark, Stefan Schulz, and Hans Åhlfeldt. In addition, we were assisted by an international team of expert reviewers, listed in the following pages.

The Scientific Programme Committee received 360 submissions of which 267 were full papers. To ensure a high quality, each submission was reviewed by at least two reviewers. The entire review process took place in the period from November 15th 2007 to February 1st 2008 and was accomplished by 210 scientific peers, who accepted 142 papers for publication. Finally, a small team of editors, consisting of Andersen, Klein, Schulz, Aarts, and Mazzoleni, reviewed the accepted papers and ensured the incorporation of the reviewers' comments and improvement suggestions.

The papers in the proceeding are grouped by themes according to the submission categories and the supplied keywords. Within the themes, the papers are in alphabetic order by the first author. As the last theme, three doctoral students from different areas of medical informatics were selected to present and discuss their research under the guidance of a panel of distinguished research faculties.

Let MIE 2008 be an opportunity to bring together all stakeholders in the eHealth to come and to exchange knowledge and insight in a lively, stimulating, and creative environment, thus creating a common ownership to the future of eHealth.

The editors would like to thank all the authors of the scientific contributions, the reviewers and our colleagues of the Scientific Programme Committee.

Stig Kjær Andersen
Gunnar O. Klein
Stefan Schulz
Jos Aarts
M. Cristina Mazzoleni

Reviewers for MIE 2008

The following persons contributed to the selection process of the papers:

Bürkle, Thomas	Chronaki, Catherine
Aarts, Jos	Chung, Grace
Abu-Hanna, Ameen	Cimino, James
Afrin, Lawrence	Costa, Carlos
Alpay, Laurence	Courtney, Karen
Alsafadi, Yasser	Criegee-Rieck, Manfred
Ammenwerth, Elske	de Bruijn, Berry
An, Ji-Young	de Keizer, Nicolette
Andersen, Stig Kjær	de la Calle, Guillermo
Angelidis, Pantelis	de Lusignan, Simon
Anguita, Alberto	Deleger, Louise
Atreja, Ashish	Demiris, George
Avillach, Paul	Denecke, Kerstin
Azevedo-Marques, Paulo	Dexheimer, Judith
Bamidis, Panagiotis	Eisenstein, Eric
Baneyx, Audrey	Elberg, Pia
Barbosa, Sayonara	Elkin, Peter
Barton, Amy	Eriksson, Henrik
Bashyam, Vijayaraghavan	Fagan, Lawrence
Bastos, Laudelino	Fernandez-Breis, Jesualdo Tomas
Benham-Hutchins, Marge	Fieschi, Marius
Bennett, Janette	Finnell, J.T.
Bernstein, Knut	Fioriglio, Gianluigi
Bertelsen, Pernille	García-Remesal, Miguel
Betts, Helen	García-Rojo, Marcial
Bittner, Thomas	Georg, Gersende
Blobel, Bernd	Georgiou, Andrew
Borycki, Elizabeth	Gibaud, Bernard
Bouaud, Jacques	Gopalan, Raj
Boyce, Richard	Ground, Anita
Boye, Niels	Hammond, Ed
Brigl, Birgit	Handels, Heinz
Brooks, Ian	Hannan, Terry
Bruun-Rasmussen, Morten	Hansen, David
Burgert, Oliver	Hansen, Margaret
Burkart, Stefan	Harkema, Henk
Bygholm, Ann	Hartvigsen, Gunnar
Charlet, Jean	Hasman, Arie
Chen, Elizabeth	Haux, Reinhold
Chen, Rong	Heinzl, Harald
Cheng, Po-Hsun	Héja, Gergely

Hellrung, Nils
 Hoehndorf, Robert
 Hofdijk, Jacob
 Horsch, Alexander
 Hristovski, Dimitar
 Hsu, William
 Hurlen, Petter
 Huser, Vojtech
 Hägglund, Maria
 Häyrinen, Kristiina
 Ingenerf, Josef
 Iyengar, Sriram
 James, Andrew
 Jao, Chiang
 Jaspers, Monique
 Jaulent, Marie-Christine
 Jenders, Robert
 Jones, Josette
 Joubert, Michel
 Joyce, Phil
 Juhola, Martti
 Kaiser, Katharina
 Kalra, Dipak
 Kastania, Anastasia
 Kaufman, David
 Kennedy, Rosemary
 Kensing, Finn
 Kern, Josipa
 Kibbe, David
 Kijsanayotin, Boonchai
 Kim, Won
 Kim, G.
 Kindler, Hauke
 Klein, Gunnar
 Knaup, Petra
 Koch, Sabine
 Kohl, Christian
 Kusnierczyk, Waclaw
 Laforest, Frederique
 Landais, Paul
 Layzell, Brian
 Le Beux, Pierre
 Liaskos, Joseph
 Lippert, Søren
 Lovis, Christian
 Luna, Daniel
 Lungeanu, Diana
 MacDonald, Ian
 Maeder, Anthony
 Magrabi, Farah
 Makikawa, Masaaki
 Malhotra, Sameer
 Manning, Bryan
 Mantas, John
 Maojo, Victor
 Marschollek, Michael
 Martín, Luis
 Martin, Suzanne
 Masic, Izet
 Mayorov, Oleg
 Mazzoleni, M. Cristina
 McCullagh, Paul
 Menke, James
 Meystre, Stéphane
 Mihalas, George
 Mittlboeck, Martina
 Moghaddam, Ramin
 Mohyuddin, Mohyuddin
 Mulvenna, Maurice
 Murray, Peter
 Mykkänen, Juha
 Müller, Marcel
 Nabeel, M.K.
 Neveol, Aurelie
 Niss, Karsten
 Nøhr, Christian
 O'Connor, Martin
 Oliveira, Jose Luis
 Onogi, Yuzo
 Oswald, Helmut
 Paiva, Paulo
 Peek, Niels
 Perez del Rey, David
 Petersson, Göran
 Petrovecki, Mladen
 Pharow, Peter
 Pinto, José Simão de Paula
 Plischke, Maik
 Popescu, Mihail
 Poprat, Michael
 Post, Andrew
 Prinz, Michael
 Punys, Vytenis
 Quaglini, Silvana
 Rahimi, Bahlol
 Rassinoux, Anne-Marie
 Rebholz-Schuhmann, Dietrich
 Renard, Jean-Marie

Richards, Janise
Roberts, Jean
Rodrigues, Jean-Marie
Roemer, Lorrie
Romacker, Martin
Rossing, Niels
Ruotsalainen, Pekka
Rydmark, Martin
Schabetsberger, Thomas
Schug, Stephan
Schulz, Stefan
Seroussi, Brigitte
Shahsavari, Nosrat
Shaw, Nicola
Shifrin, Michael
Smalheiser, Neil
Smith, Catherine
Splendiani, Andrea
Spyns, Peter
Srinivasan, Arunkumar
Staemmler, Martin
Stamatopoulos, Vasileios
Stenzhorn, Holger
Stern, Milton
Stoicu-Tivadar, Lacramioara
Supek, Selma
Surján, György
Svanæs, Dag
Szabo, Zoltan
Tachinardi, Umberto
Talmon, Jan
Thurin, Anders
Timpka, Toomas
Tromp, Miranda
Tsafnat, Guy
Tsiknakis, Manolis
Tusch, Guenter
Vimarlund, Vivian
Vizenor, Lowell
Weber, Patrick
Weltner, János
Westbrook, Johanna
Wolf, Klaus-Hendrik
Wright, Adam
Yen, Po-Yin
Zai, Adrian
Zhang, Songmao
Zhou, Li
Zweigenbaum, Pierre
Åhlfeldt, Hans

This page intentionally left blank

Contents

Preface	v
<i>Stig Kjær Andersen, Gunnar O. Klein, Stefan Schulz, Jos Aarts and M. Cristina Mazzoleni</i>	
Reviewers for MIE 2008	vii
1. Bioinformatics	
An Ontology Based Method to Solve Query Identifier Heterogeneity in Post-Genomic Clinical Trials	3
<i>Alberto Anguita, Luis Martín, José Crespo and Manolis Tsiknakis</i>	
Gene Regulation Ontology (GRO): Design Principles and Use Cases	9
<i>Elena Beisswanger, Vivian Lee, Jung-Jae Kim, Dietrich Rebholz-Schuhmann, Andrea Splendiani, Olivier Dameron, Stefan Schulz and Udo Hahn</i>	
Differences in Doctors' and Nurses' Assessments of Hospital Culture and Their Views About Computerised Order Entry Systems	15
<i>Joanne Callen, Jeffrey Braithwaite and Johanna Westbrook</i>	
Novelty Detection Using One-Class Parzen Density Estimator. An Application to Surveillance of Nosocomial Infections	21
<i>Gilles Cohen, Hugo Sax and Antoine Geissbuhler</i>	
Interface Terminologies: Bridging the Gap Between Theory and Reality for Africa	27
<i>Andrew S. Kanter, Amy Y. Wang, Fred E. Masarie, Frank Naeymi-Rad and Charles Safran</i>	
Cataloguing and Displaying Web Feeds from French Language Health Sites: A Web 2.0 Add-on to a Health Gateway	33
<i>G. Kerdelhué, B. Thirion, B. Dahamna and S.J. Darmoni</i>	
Videophones for the Delivery of Home Healthcare in Oncology	39
<i>Mona Laila, Vincent Rialle, Lydie Nicolas, Catherine Duguay and Alain Franco</i>	
Addressing the Biomedical Informatics Needs of a Microarray Laboratory in a Clinical Microbiology Context	45
<i>Guillermo Lopez-Campos, Victoria Lopez Alonso and Fernando Martin-Sanchez</i>	
CEMARA: A Web Dynamic Application Within a N-Tier Architecture for Rare Diseases	51
<i>Claude Messiaen, Loïc Le Mignot, Ana Rath, Jean-Baptiste Richard, Eric Dufour, Mohamed Ben Said, Jean-Philippe Jais, Alain Verloes, Martine Le Merrer, Christine Bodemer, Geneviève Baujat, Marion Gerard-Blanluet, Eva Bourdon-Lanoy, Rémi Salomon, Ségolène Ayme and Paul Landais for the CEMARA Task Force</i>	
Developing a Shared Electronic Health Record for Patients and Clinicians	57
<i>Cornelia M. Ruland, Haakon Brynhi, Roar Andersen and Tore Bryhni</i>	
Examination of Computer Assisted Prescribing of an Initial Calculated Antibiotic Treatment	63
<i>R. Röhrig, E.J. Niczko, H. Beuteführ, S. Böttger, J. Klasen, R. Füssle and B. Hartmann</i>	

2. Consumer and Home Based eHealth

Treasure Hunt – A Serious Game to Support Psychotherapeutic Treatment of Children	71
<i>Veronika Brezinka</i>	
Virtual Rehabilitation After Stroke	77
<i>Jurgen Broeren, Ann Bjorkdahl, Lisbeth Claesson, Daniel Goude, Åsa Lundgren-Nilsson, Hans Samuelsson, Christian Blomstrand, Katharina S. Sunnerhagen and Martin Rydmark</i>	
An Easy to Use and Affordable Home-Based Personal eHealth System for Chronic Disease Management Based on Free Open Source Software	83
<i>Tatjana M. Burkow, Lars K. Vognild, Trine Krogstad, Njål Borch, Geir Østengen, Astrid Bratvold and Marijke Jongsma Risberg</i>	
Aligning Lay and Specialized Passages in Comparable Medical Corpora	89
<i>Louise Deleger and Pierre Zweigenbaum</i>	
Knowledge Engineering as a Support for Building an Actor Profile Ontology for Integrating Home-Care Systems	95
<i>Karina Gibert, Aida Valls and David Riaño</i>	
Evaluation of the Use of Digital Pens for Pain Assessment in Palliative Home Healthcare	101
<i>Leili Lind</i>	
Reusing Models of Actors and Services in Smart Homecare to Improve Sustainability	107
<i>Ståle Walderhaug, Erlend Stav and Marius Mikalsen</i>	
A System for Monitoring Physical Activity Data Among People with Type 2 Diabetes	113
<i>Eirik Årsand, Odd-Arne Olsen, Ragnhild Varmedal, Willy Mortensen and Gunnar Hartvigsen</i>	

3. Decision Support and Knowledge Management

Medical Knowledge Packages and Their Integration into Health-Care Information Systems and the World Wide Web	121
<i>Klaus-Peter Adlassnig and Andrea Rappelsberger</i>	
Diversity in Preoperative-Assessment Data Collection, a Literature Review	127
<i>Leila Ahmadian, Ronald Cornet, Wilton A. van Klei and Nicolette F. de Keizer</i>	
Representing Health, Disorder and Their Transitions by Digraphs	133
<i>Carl-Fredrik Bassøe</i>	
Characterizing the Dimensions of Clinical Practice Guideline Evolution	139
<i>Jacques Bouaud and Brigitte Séroussi</i>	
Specification of Business Rules for the Development of Hospital Alarm System: Application to the Pharmaceutical Validation	145
<i>Abdelali Boussadi, Cedric Bousquet, Brigitte Sabatier, Isabelle Colombet and Patrice Degoulet</i>	
Formalization of Clinical Practice Guidelines	151
<i>David Buchtela, Jan Peleška, Arnošt Veselý, Jana Zvárová and Miroslav Zvolský</i>	

A Simple Method for Heuristic Modeling of Expert Knowledge in Chronic Disease: Identification of Prognostic Subgroups in Rheumatology <i>Örjan Dahlström, Toomas Timpka, Ursula Hass, Thomas Skogh and Ingrid Thyberg</i>	157
A Method for Indexing Biomedical Resources over the Internet <i>Guillermo de la Calle, Miguel Garcia-Remesal and Victor Maojo</i>	163
Improving Pain & Symptom Management for Advanced Cancer Patients with a Clinical Decision Support System <i>Krista Elvidge</i>	169
Collaboration Patterns in an Online Community of Practice in Oral Medicine <i>Göran Falkman, Marie Gustafsson, Olof Torgersson and Mats Jontell</i>	175
An Automated Personalised Intervention Algorithm for Remote Patient Monitoring <i>Joanna Fursse, Malcolm Clarke, Russell Jones, Sneh Khemka and Genevieve Findlay</i>	181
Syntactical Negation Detection in Clinical Practice Guidelines <i>Stefan Gindl, Katharina Kaiser and Silvia Miksch</i>	187
Subjective Usability of the CARDSS Guideline-Based Decision Support System <i>Rick Goud, Monique W.M. Jaspers, Arie Hasman and Niels Peek</i>	193
Assessment of Biomedical Knowledge According to Confidence Criteria <i>Ines Jilani, Natalia Grabar, Pierre Meneton and Marie-Christine Jaulent</i>	199
Using Knowledge for Indexing Health Web Resources in a Quality-Controlled Gateway <i>Michel Joubert, Stefan J. Darmoni, Paul Avillach, Badisse Dahamna and Marius Fieschi</i>	205
On Machine Learning Classification of Otoneurological Data <i>Martti Juhola</i>	211
Semantic Relation Mining of Solid Compounds in Medical Corpora <i>Dimitrios Kokkinakis</i>	217
Use of the C4.5 Machine Learning Algorithm to Test a Clinical Guideline-Based Decision Support System <i>Jean-Baptiste Lamy, Anis Ellini, Vahid Ebrahimi, Jean-Daniel Zucker, Hector Falcoff and Alain Venot</i>	223
Process Mining for Clinical Workflows: Challenges and Current Limitations <i>Martin Lang, Thomas Bürkle, Susanne Laumann and Hans-Ulrich Prokosch</i>	229
Searching Related Resources in a Quality Controlled Health Gateway: A Feasibility Study <i>Tayeb Merabti, Suzanne Pereira, Catherine Letord, Thierry Lecroq, Michel Joubert, Badisse Dahamna and Stéfan J. Darmoni</i>	235
Association Studies on Cervical Cancer Facilitated by Inference and Semantic Technologies: The ASSIST Approach <i>Pericles Mitkas, Vassilis Koutkias, Andreas Symeonidis, Manolis Falelakis, Christos Diou, Irimi Lekka, Anastasios Delopoulos, Theodoros Agorastos and Nicos Maglaveras</i>	241
Discrete Event Simulation as a Tool in Optimization of a Professional Complex Adaptive System <i>Anders Lassen Nielsen, Helmer Hilwig, Niranjana Kissoon and Surujpal Teelucksingh</i>	247

An Artificial Neural Network Derived Trauma Outcome Prediction Score as an Aid to Triage for Non-Clinicians <i>Adrian Pearl, Raphael Bar-Or and David Bar-Or</i>	253
Case-Based Reasoning to Explain Medical Model Exceptions <i>Rainer Schmidt and Olga Vorobieva</i>	259
Diagnostic Games: From Adequate Formalization of Clinical Experience to Structure Discovery <i>Michael A. Shifrin and Eva I. Kasparova</i>	265
Disease Outbreak Detection Through Clique Covering on a Weighted ICPC-Coded Graph <i>Klaske van Vuurden, Gunnar Hartvigsen and Johan Gustav Bellika</i>	271
4. Evaluation	
A Randomized Controlled Trial to Evaluate an Electronic Scoring Tool in the ICU <i>Thomas Birkle, Alexander Beisig, Marion Ganslmayer and Hans-Ulrich Prokosch</i>	279
Evaluating the Impact of a Service-Oriented Framework for Healthcare Interoperability <i>Stylianos Daskalakis and John Mantas</i>	285
The Use of Performance Metrics to Monitor the Impact of CPOE on Pathology Laboratory Services <i>Andrew Georgiou, Wendy Morse, Wyndham Timmins, Sangeeta Ray and Johanna I. Westbrook</i>	291
Development of a Patient-Oriented Tool for Evaluating the Quality of Breast Cancer Information on the Internet <i>Wen-Chin Hsu and Peter A. Bath</i>	297
Pre-Post Evaluation of Physicians' Satisfaction with a Redesigned Electronic Medical Record System <i>Monique W.M. Jaspers, Linda W.P. Peute, Arnaud Lauteslager and Piet J.M. Bakker</i>	303
CPOE System Design Aspects and Their Qualitative Effect on Usability <i>Reza Khajouei and Monique W.M. Jaspers</i>	309
Decision Support System Supporting Clinical Reasoning Process – An Evaluation Study in Dementia Care <i>Helena Lindgren</i>	315
Evaluating Inter-Professional Work Support by a Computerized Physician Order Entry (CPOE) System <i>Zahra Niazkhani, Habibollah Pirnejad, Antoinette de Bont and Jos Aarts</i>	321
Usability Studies on Interactive Health Information Systems; Where Do We Stand? <i>Linda W.P. Peute, Richard Spithoven, Piet J.M. Bakker and Monique W.M. Jaspers</i>	327
Combination of Short- and Longaxis MR Image Sequences for the 3D Segmentation of the Left Ventricle <i>Dennis Säring, Jatin Relan, Michael Groth, Kai Müllerleile and Heinz Handels</i>	333

An Automated Method for Analyzing Adherence to Therapeutic Guidelines: Application in Diabetes	339
<i>Massoud Toussi, Vahid Ebrahimi, Philippe Le Toumelin, Régis Cohen and Alain Venot</i>	
Computerised Order Entry Systems: Sustained Impact on Laboratory Efficiency and Mortality Rates?	345
<i>Johanna I. Westbrook, Andrew Georgiou and Marilyn I. Rob</i>	
5. Health Information Systems Including EHR	
Managing Care Pathways Combining SNOMED CT, Archetypes and an Electronic Guideline System	353
<i>Knut Bernstein and Ulrich Andersen</i>	
Analysis and Evaluation of EHR Approaches	359
<i>Bernd G.M.E. Blobel and Peter Pharow</i>	
Electronic Disease Surveillance for Sensitive Population Groups – The Diabetics Case Study	365
<i>Taxiarchis Botsis, Ole Hejlesen, Johan Gustav Bellika and Gunnar Hartvigsen</i>	
Clinical Processes in an Innovative Vascular Surgeon Community. Implications for Workflow Modeling	371
<i>Berit Brattheim, Andreas R. Seim and Arild Faxvaag</i>	
Medical Knowledge Representation System	377
<i>David Buchtela, Jan Peleska, Miroslav Zvolsky and Jana Zvarova</i>	
An Electronic Registry for Physiotherapists in Belgium	383
<i>Ronald Buyl and Marc Nyssen</i>	
Computer Support for Shared Care of Diabetes: Findings from a Danish Case	389
<i>Keld Bødker and Maren Fich Granlien</i>	
From Documents on Paper to Electronic Medical Records	395
<i>Lino Carrajo, Ángel Penas, Rubén Melcón, Fco. Javier González and Eduardo Couto</i>	
An Archetype-Based Testing Framework	401
<i>Rong Chen, Sebastian Garde, Thomas Beale, Mikael Nyström, Daniel Karlsson, Gunnar O. Klein and Hans Åhlfeldt</i>	
Supervised Approach to Recognize Question Type in a QA System for Health	407
<i>Sarah Cruchet, Arnaud Gaudinat and Célia Boyer</i>	
Group Decision Support System Applied to the Medical Pluri-Disciplinary Decision Group: Usability and Efficacy	413
<i>Nathalie Degardin-Capon, Nathalie Bricon-Souf, Marie-Catherine Beuscart-Zephir and Régis Beuscart</i>	
The Gap Between Actual and Mandated Use of an Electronic Medication Record Three Years After Deployment	419
<i>Maren Fich Granlien, Morten Hertzum and Jette Gudmundsen</i>	
Analysis of EHRs for Research, Quality Management and Health Politics	425
<i>Walter Gall, Wilfried Grossmann, Georg Dufts Schmid, Thomas Wrba and Wolfgang Dorda</i>	
Consent-Based Access to Core EHR Information: The SUMO-Project	431
<i>Vigdis Heimly</i>	

Facilitating the openEHR Approach – Organizational Structures for Defining High-Quality Archetypes	437
<i>Christian Dominik Kohl, Sebastian Garde and Petra Knaup</i>	
Implementation of an Electronic Medication System and Disregarded Power of the Record	443
<i>Henriette Mabeck</i>	
cyberMarathon – Increasing Physical Activity Using Health-Enabling Technologies	449
<i>Maik Plischke, Michael Marscholke, Klaus-Hendrik Wolf, Reinhold Haux and Uwe Tegtbur</i>	
Integrating Clinical, Gene Expression, Protein Expression and Preanalytical Data for <i>in silico</i> Cancer Research	455
<i>Delphine Rossille, Anita Burgun, Céline Pangault-Lorho and Thierry Fest</i>	
Developing a Taxonomy of Communication Errors in Heterogeneous Information Systems	461
<i>Samrend Saboor and Elske Ammenwerth</i>	
Mining Knowledge from Corpora: An Application to Retrieval and Indexing	467
<i>Lina F. Soualmia, Badisse Dahamna and Stéfan Darmoni</i>	
Enhanced Information Retrieval from Narrative German-Language Clinical Text Documents Using Automated Document Classification	473
<i>Stephan Spat, Bruno Cadonna, Ivo Rakovac, Christian Gütl, Hubert Leitner, Günther Stark and Peter Beck</i>	
Frequency of Hospital-Acquired Pneumonia in Electronic and Paper-Based Patient Record	479
<i>Jürgen Stausberg and Abdelouahid Azaouagh</i>	
Reliable Personal Health Records	484
<i>Ton van Deursen, Paul Koster and Milan Petković</i>	
Involving Clinicians in the Development of an Electronic Clinical Handover System – Thinking Systems Not Just Technology	490
<i>Ming Chao Wong, Paul Turner and Kwang Chien Yee</i>	
Perfect Match? Generation Y as Change Agents for Information Communication Technology Implementation in Healthcare	496
<i>Kwang Chien Yee, Erin Miils and Caroline Airey</i>	
6. Human-Computer Interaction & Imaging	
Does a Hybrid Electronic-Paper Environment Impact on Health Professional Information Seeking?	505
<i>E.M. Borycki and L. Lemieux-Charles</i>	
Application of Business Process Management to Drive the Deployment of a Speech Recognition System in a Healthcare Organization	511
<i>Maria José González Sánchez, José Manuel Framiñán Torres, Carlos Luis Parra Calderón, Juan Antonio Del Rio Ortega, Eduardo Vigil Martín and Jaime Nieto Cervera</i>	
Affective Computing and Medical Informatics: State of the Art in Emotion-Aware Medical Applications	517
<i>Andrej Lunski, Panagiotis D. Bamidis and Madga Hitoglou-Antoniadou</i>	
Using Medline Queries to Generate Image Retrieval Tasks for Benchmarking	523
<i>Henning Müller, Jayashree Kalpathy-Cramer, William Hersh and Antoine Geissbuhler</i>	

Voice-Controlled Data Entry in Dental Electronic Health Record <i>Miroslav Nagy, Petr Hanzlicek, Jana Zvarova, Tatjana Dostalova, Michaela Seydlova, Radim Hippman, Lubos Smidl, Jan Trmal and Josef Psutka</i>	529
A Large, High Resolution Tiled Display for Medical Use: Experiences from Prototyping of a Radiology Scenario <i>Bernt Ivar Olsen, Sanjaya Babu Dhakal, Odd Petter Eldevik, Per Hasvold and Gunnar Hartvigsen</i>	535
The Contextual Nature of Usability and Its Relevance to Medical Informatics <i>Dag Svanæs, Anita Das and Ole Andreas Alsos</i>	541
OPTISAS a New Method to Analyse Patients with Sleep Apnea Syndrome <i>Adrien Ugon, Carole Philippe, Slawomir Pietrasz, Jean-Gabriel Ganascia and Pierre P. Levy</i>	547
7. Learning, Modelling and Simulation	
Predictors of Preterm Birth in Birth Certificate Data <i>Karen L. Courtney, Sara Stewart, Mihail Popescu and Linda K. Goodwin</i>	555
The p53 Network Modeling – Current State and Future Prospects <i>Raul Florin Horhat, Gheorghe I. Mihalas and Mihaela Neamtu</i>	561
Using a Low-Cost Simulation Approach for Assessing the Impact of a Medication Administration System on Workflow <i>A.W. Kushniruk, E.M. Borycki, S. Kuwata and H. Watanabe</i>	567
Process Mining Techniques: An Application to Stroke Care <i>Ronny Mans, Helen Schonenberg, Giorgio Leonardi, Silvia Panzarasa, Anna Cavallini, Silvana Quaglini and Wil van der Aalst</i>	573
Development of an E-Learning System for Occupational Medicine: Usability Issues <i>M. Cristina Mazzoleni, Carla Rognoni, Enrico Finozzi, Ines Giorgi, Franco Pugliese, Marco Pagani and Marcello Imbriani</i>	579
Economic Advantage of Pharmacogenomics – Clinical Trials with Genetic Information <i>Wataru Ohashi, Hiroshi Mizushima and Hiroshi Tanaka</i>	585
Recognising e-Health as Part of a Cohesive Professional Community <i>Jean M. Roberts</i>	591
Application of the Multi-Disciplinary Thematic Seminar Method in Two Homecare Cases – A Comparative Study <i>Isabella Scandurra, Maria Hägglund and Sabine Koch</i>	597
8. National eHealth Roadmaps, Cross-Border Applications and Organisational Strategies	
Modelling Access to Renal Transplantation Waiting List in a French Healthcare Network Using a Bayesian Method <i>Sahar Bayat, Marc Cuggia, Michel Kessler, Serge Briançon, Pierre Le Beux and Luc Frimat</i>	605
Evaluation of Robustness of a User Requirements Specification Approach in a Purchase Context, a LIS Case Study <i>Jytte Brender and Peter McNair</i>	611
A Vision for the Use of ICT by Norwegian Healthcare in 2012 <i>Edgar Glück, Roald Bergström, Ole-Fredrik Melleby and Jessica Gabin</i>	617

IT-Based Information Management in Health Care Networks: The MedoCom Approach	623
<i>Nils Hellrung, Nathalie Gusew, Martin Willkomm and Reinhold Haux</i>	
The Role of Patients and Their Health Cards in Integrated eHealth Environments	629
<i>Claudia Hildebrand, Peter Pharow and Bernd Blobel</i>	
A Deployment and Research Roadmap for Semantic Interoperability: The EU SemanticHEALTH Project	635
<i>Pierre Lewalle, Jean M. Rodrigues, Pieter Zanstra, Bedirhan Ustun, Dipak Kalra, Gyorgy Surjan, Alan Rector, Veli Stroetmann and Martti Virtanen</i>	
DebugIT for Patient Safety – Improving the Treatment with Antibiotics Through Multimedia Data Mining of Heterogeneous Clinical Data	641
<i>Christian Lovis, Dirk Colaert and Veli N. Stroetmann</i>	
Study on Urban Healthcare Consumption in Northern France	647
<i>Julia Salleron, Alain Duhamel, Charles Sulman, Valérie Deken and Régis Beuscart</i>	
Cross-Border Collaboration Between Greece and FYROM: Mobile Healthcare Provision	653
<i>Stergiani Spyrou, Dimitrios Vartzopoulos, Panagiotis Bamidis and Nicos Maglaveras</i>	
9. Privacy and Security	
Secure Dissemination of Electronic Healthcare Records in Distributed Wireless Environments	661
<i>Petros Belsis, Dimitris Vassis, Christos Skourlas and Grammati Pantziou</i>	
Watermarking Medical Images with Anonymous Patient Identification to Verify Authenticity	667
<i>Gouenou Coatrieux, Catherine Quantin, Julien Montagner, Maniane Fassa, François-André Allaert and Christian Roux</i>	
Empowerment of Health Professionals: How High Level Security Education Can Raise Awareness and Confidence	673
<i>Matthias Herbst, Christoph Busch, Peter Pharow and Bernd Blobel</i>	
Context-Aware Access Control for Pervasive Access to Process-Based Healthcare Systems	679
<i>Vassiliki Koufi and George Vassilacopoulos</i>	
Knowledge Management for the Protection of Information in Electronic Medical Records	685
<i>Nathan Lea, Stephen Hailes, Tony Austin and Dipak Kalra</i>	
Improving Patients Privacy with Pseudonymization	691
<i>Thomas Neubauer and Bernhard Riedl</i>	
Mobile Health Requires Mobile Security: Challenges, Solutions, and Standardization	697
<i>Peter Pharow and Bernd Blobel</i>	
Towards Dynamic Access Control for Healthcare Information Systems	703
<i>Lillian Røstad and Øystein Nytrø</i>	
Security, Safety, and Related Technology – The Triangle of eHealth Service Provision	709
<i>Mario Savastano, Ashbjorn Hovsto, Peter Pharow and Bernd Blobel</i>	

10. Standardization

Developing a Standard for Personal Health Devices Based on 11073 <i>Malcolm Clarke</i>	717
SOA Approach for Integration of Departmental Systems <i>Timo Itälä, Jari Ukkola, Aino Virtanen and Juha Mykkänen</i>	723
Procurement of Prescriber Support Systems <i>Karin Kajbjer</i>	729
Enhanced Semantic Interpretability by HealthCare Standards Profiling <i>Diego M. Lopez and Bernd G.M.E. Blobel</i>	735
Reconciling Data Structures in Health Information Systems for Research in a European Clinical Registry <i>Thomas H. Müller</i>	741
Using ESB and BPEL for Evolving Healthcare Systems Towards SOA <i>D. Papakonstantinou, F. Malamateniou and G. Vassilacopoulos</i>	747
Systematizing Medical Alerts <i>Mattias Pettersson, Jenny Wihlborg, Rikard Löfvström, Erik Sundvall, Mikael Nyström and Daniel Karlsson</i>	753
A Framework for Semantic Interoperability in Healthcare: A Service Oriented Architecture Based on Health Informatics Standards <i>Amanda Ryan and Peter Eklund</i>	759
The Adoption of IT Security Standards in a Healthcare Environment <i>Rui Gomes and Luís Velez Lapão</i>	765
IHE Based Interoperability – Benefits and Challenges <i>Florian Wozak, Elske Ammenwerth, Alexander Hörbst, Peter Sögner, Richard Mair and Thomas Schabetsberger</i>	771

11. Terminology and Ontology

Cross-Mapping APACHE IV “Reasons for Intensive Care Admission” Classification to SNOMED CT <i>F. Bakhshi-Raiez, R. Cornet and N.F. de Keizer</i>	779
Do SNOMED CT Relationships Qualify? <i>Ronald Cornet</i>	785
Enhancing Knowledge Representations by Ontological Relations <i>Kerstin Denecke</i>	791
Biosurveillance Evaluation of SNOMED CT’s Terminology (BEST Trial): Coverage of Chief Complaints <i>Peter L. Elkin, Steven H. Brown, Andrew Balas, Zelalem Temesgen, Dietlind Wahner-Roedler, David Froehling, Mark Liebow, Brett Trusko, S. Trent Rosenbloom and Greg Poland</i>	797
Mining for Adverse Drug Events with Formal Concept Analysis <i>Alexander Estacio-Moreno, Yannick Toussaint and Cédric Bousquet</i>	803
Automatic Acquisition of Synonyms from French UMLS for Enhanced Search of EHRs <i>Natalia Grabar, Paul-Christophe Varoutas, Philippe Rizand, Alain Livartowski and Thierry Hamon</i>	809
Design of an Automatic Coding Algorithm for a Multi-Axial Classification in Pathology <i>André Happe, Marc Cuggia, Bruno Turlin and Pierre Le Beux</i>	815

Design Principles of DOLCE-Based Formal Representation of ICD10 <i>Gergely Héja, Péter Varga and György Surján</i>	821
A Version Management System for SNOMED CT <i>Josef Ingenerf and Thomas Beisiegel</i>	827
Representation of Disorders of the Newborn Infant by SNOMED CT® <i>Andrew G. James and Kent A. Spackman</i>	833
Exploratory Analysis of Medical Coding Practices: The Relevance of Reported Data Quality in Obstetrics-Gynaecology <i>Diana Lungeanu, Daniela Zaharie, Stefan Holban, Elena Bernad, Maria Bari and Rodica Noaghiu</i>	839
Use of Super-Concepts to Customize Electronic Medical Records Data Display <i>P. Massari, S. Pereira, B. Thirion, A. Derville and S.J. Darmoni</i>	845
Semantic Web Ontology Utilization for Heart Failure Expert System Design <i>Marin Prcela, Dragan Gamberger and Alan Jovic</i>	851
Standards and Biomedical Terminologies: The CEN TC 251 and ISO TC 215 Categorical Structures. A Step Towards Increased Interoperability <i>Jean M. Rodrigues, Anand Kumar, Cédric Bousquet and Béatrice Trombert</i>	857
How Granularity Issues Concern Biomedical Ontology Integration <i>Stefan Schulz, Martin Boeker and Holger Stenzhorn</i>	863
Dichotomy – A Forgotten Ancient Principle <i>György Surján</i>	869
12. Doctoral Consortia Papers	
Attention and Usability Issues in Mobile Health Information Systems at Point-of-Care <i>Ole Andreas Alsos</i>	877
Reconstructing Clinical Events by Interpreting NICU Monitoring Data <i>Feng Gao</i>	879
Evaluating the Impact of CPOE Systems on Medical Workflow: A Mixed Method Study <i>Zahra Niazkhani</i>	881
Author Index	883

1. Bioinformatics

This page intentionally left blank

An Ontology Based Method to Solve Query Identifier Heterogeneity in Post-Genomic Clinical Trials

Alberto ANGUITA^a, Luis MARTÍN^a, José CRESPO^a and Manolis TSIKNAKIS^b

^a*Biomedical Informatics Group, Artificial Intelligence Laboratory,*

School of Computer Science, Universidad Politécnica de Madrid

Campus de Montegancedo S/N, 28660 Boadilla del Monte, Madrid, Spain

^b*Center of eHealth Technologies, Institute of Computer Science,*

Foundation for Research and Technology - Hellas

GR-71110 Heraklion, Crete, Greece

Abstract. The increasing amount of information available for biomedical research has led to issues related to knowledge discovery in large collections of data. Moreover, Information Retrieval techniques must consider heterogeneities present in databases, initially belonging to different domains—e.g. clinical and genetic data. One of the goals, among others, of the ACGT European is to provide seamless and homogeneous access to integrated databases. In this work, we describe an approach to overcome heterogeneities in identifiers inside queries. We present an ontology classifying the most common identifier semantic heterogeneities, and a service that makes use of it to cope with the problem using the described approach. Finally, we illustrate the solution by analysing a set of real queries.

Keywords. Ontologies, Clinical Trials, Database Integration, Semantic Mediation

Introduction

Advancing Clinico-Genomic Trials on Cancer (ACGT) is a European Commission-supported project that aims to develop a set of ontology-driven technologies to support the development of treatment and research within post-genomic multicentric clinical trials on cancer. One of the main issues that ACGT is trying to tackle is providing seamless, homogeneous syntactic and semantic access to a single virtual repository representing the set of integrated databases needed within a specific clinical trial. To achieve this goal, a data access infrastructure is being developed. This infrastructure is comprised by several tools: the ACGT Master Ontology on Cancer—an ontology modelling the domain of clinical trials on cancer—, the ACGT Data Access Services—coping with syntactic heterogeneities—and the ACGT Semantic Mediator—providing database integration and semantic homogenization. In order to facilitate the communication between different services, XML has been selected as the information representation format: the Master Ontology is developed in OWL, database schemas

are represented in RDFS by the Data Access Services and the adopted query language is SPARQL.

The ACGT Semantic Mediator deals with a variety of problems related to data semantic heterogeneities. These heterogeneities can be classified in two categories: i) Schema Level Heterogeneities, and ii) Instance/identifier Level heterogeneities. While the former are treated by mapping the RDFS representing the database schema information using the Master Ontology as a semantic framework, the latter cannot be tackled using this model. During the normal execution of a query through the Semantic Mediator, instance level heterogeneities are present in two different places: i) the retrieved results, and ii) identifiers in queries. We have proposed two approaches to address these issues. The first one is called *OntoDataClean* [1], and uses an ontology that represents the most common cases of instance level heterogeneity in the biomedical data. In this paper, we present the second approach, to overcome instance level heterogeneities in queries. It is based on the same principle proposed on *OntoDataClean*: using a transversal ontology representing the domain of instance level heterogeneities to produce a query translation overcoming these issues.

This paper is organized as follows: first, we give an overview of the state of the art and related work on semantic mediation and the study of heterogeneities. Then, we describe our approach and tool to overcome query identifier heterogeneities. In the case study section we show how our approach behaves with several real queries. Finally, we conclude summarizing our experiences so far.

1. Background & Related Work

The problem of database integration has been intensely studied during the last 20 years. We can classify the main approaches in three categories: i) Data Translation, ii) Query Translation, and iii) Information Linkage. Data Translation approaches are based in the actual homogenization of data for its subsequent storage in a central repository. Conversely, Query Translation approaches leave the data in the original databases, and expose a view to the user expressing the set of possible queries. When a query is launched, the mediation software splits and translates it into dedicated queries for the underlying databases. By contrast, Information Linkage approaches maintain a set of links among sources using cross reference. Both Data Translation and Query Translation approaches must take into consideration the possible semantic heterogeneities, but only the latter need to process queries in order to overcome them.

Formal semantics have formed part of the main efforts carried out in the fields of Knowledge Recovery and Database Management during the last years [2]. According to [3], XML can be used to overcome syntactic heterogeneities, while extensions of it, such as OWL introduce the semantic models as descriptions of domains—i.e. ontologies. However, semantic problems such as dealing with different descriptions of a single concept must be solved, even when ontologies are used as homogenization framework.

The classical role of ontologies in database integration is acting as a general description of the domain the different data belong to —e.g. clinical, genetic and image data, in the case of ACGT. Domain information from the ontology is mapped to terms and relations belonging to the databases. This mapping information will be used to translate a query or the actual data, depending on the selected approach. This kind of usage of ontologies has obtained success in different projects, such as ONTOFUSION

[4], SEMEDA [5], KAON Reverse [6] and D2RMAP [7]. These kinds of approaches treat mainly schema level heterogeneities.

By contrast, less amount of effort has been expended in resolving heterogeneities in the actual data—i.e. instance level heterogeneities. An example of tool that makes use of ontologies to overcome instance level heterogeneities in data is *OntoDataClean* [1]. This tool performs data processing in the results retrieving process, using an ontology that classifies the most common cases of semantic heterogeneity to identify and process data. To the best of our knowledge, no automatic ontology-based method to process identifier query heterogeneities exists. Some interesting studies on the types of identifier heterogeneities can be found, such as the classification given by [8].

2. Method

There exist different types of semantic heterogeneities that can be found when dealing with data coming from different sources, maybe belonging to heterogeneous domains. When formulating a query, a user can express the constraints of the view she wants to be retrieved by using explicit literals as values for determined fields within a database. A single database can present heterogeneity in its contained data—i.e. mainly because of bad information management before the storage, or the use of imprecise tools to collect them—, but the problems become greater when dealing with integrated databases. In the ideal case, a user that needs to launch a query does not want to be concerned about possible heterogeneities in the underlying data. Moreover, she may want to be given a standard way to express both terms and literals in her query—e.g. data can present scale representation heterogeneities, such as number of white blood cells per mililiter or microliter, depending on the laboratory that did the blood analysis. The question is the following: How a clinician should express a constraint in a query? The answer is simple: the system should give the user the view that there is only one way, which is the selected standard, and should be able to tackle deviations from this standard in the queries if they are presented. The mediation system must undertake then the responsibility of translating not only the terms, but the literals expressed in the query so that proper subqueries are sent to the underlying databases.

We propose an ontology-based approach, *OntoQueryClean*, to overcome these kinds of heterogeneities. *OntoQueryClean* is the evolution of *OntoDataClean*, a tool for data preprocessing also based on ontologies, and which forms part of the *ONTOFUSION* system for integration of distributed and heterogeneous data sources. *OntoDataClean* uses an ontology that proposes a classification of possible heterogeneities, and that defines a set of basic transformation methods. Instances of this preprocessing ontology can be created in order to specify how data have to be modified, allowing proper integration of heterogeneous sources. In *OntoQueryClean*, a new preprocessing ontology has been built, based on query preprocessing requirements, describing the most common heterogeneities that can be found. Three transformation methods are defined, namely i) Scale transformations, ii) Format transformations and iii) Synonym transformations. Scale transformations allow solving heterogeneities due to the use of different scales among a set of sources—i.e. by defining arithmetic expressions that are applied over numeric data—, such as the previous example regarding measurements of white blood cells. Format transformations permit modifying string values through either regular expressions or rule algorithms. With these transformations we can, for example, modify the format of a date from MM-DD-

YYYY to DD/MM/YY. Synonym transformations allow defining pairs of synonyms so the system performs adequate substitutions.

Prior to actual query cleaning, the literals inside a specific query must be properly identified—i.e. recognize which concept they are instantiating. For this, a specific purpose module has been developed. This module is able to parse SPARQL queries and RDF Schema files to associate each literal included in the given query with a class in the schema. Figure 1 depicts the architecture of OntoQueryClean.

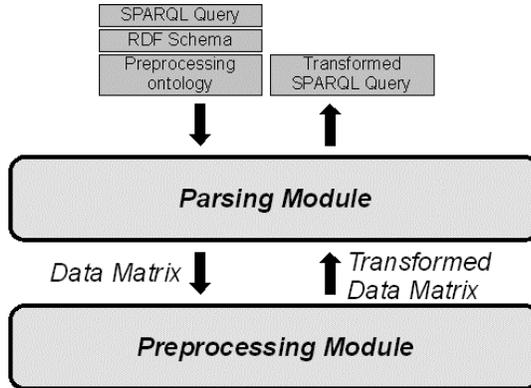


Figure 1. Architecture of the OntoQueryClean system

OntoQueryClean is part of the ACGT Semantic Mediator, taking care of the query preprocessing. Whenever a query is launched against the Mediator, this must create equivalent subqueries for the underlying sources. Once these are generated, they are given to the OntoQueryClean tool so literals are properly transformed and follow the format imposed by each database. Previous work has to be carried out in order to define preprocessing ontologies for each source. This task can be done using any existing ontology editor, such as Protégé or SWOOP.

3. Results

We have tested our tools by conducting experiments using three different data sources, from two clinical trials ACGT is working in—i.e. TOP and SIOP—and a DICOM repository of images. Test versions of these databases were incorporated in the ACGT Semantic Mediator infrastructure, so integrated queries could be performed. As stated before, OntoQueryClean is accessed once the schema level heterogeneities have been already tackled by the Mediator. A set of integrated queries was launched against the Mediator, and thus preprocessed by OntoQueryClean. In the following paragraphs we describe two of these queries, showing how the tool solved the heterogeneities found.

The first example shows a synonym transformation applied over a literal indicating a patient identifier. A global query asking for hospital identifiers where a specific patient was treated and clinical study identifiers for that same patient is presented. This query involves only the SIOP and DICOM databases, thus only queries for these are generated. The identifier used in the global query corresponds to the existing one in SIOP. However, it differs from the identifiers contained in DICOM. Synonym

transformations are carried out in order to convert the string identifiers into the format used in DICOM. Figure 2 shows the queries before and after translation.

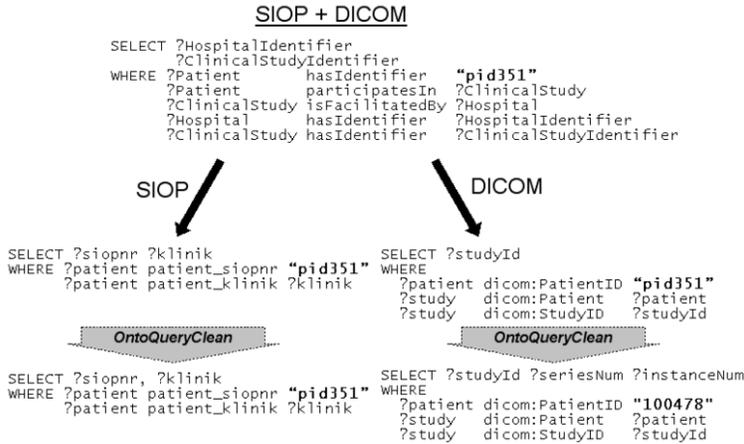


Figure 2. Synonym transformations over an integrated query for SIOP + DICOM

As can be seen, a general query expressed in terms of the global schema of the integrated repository is launched. This query includes a constraint on the patient identifier. The Mediator undertakes the task of splitting this query into dedicated subqueries for both SIOP and DICOM. However, the Mediator itself does not deal with the possible heterogeneity in the literal included in the constraint. In order to cope with this issue, *OntoQueryClean* is invoked. In the case of the SIOP query, *OntoQueryClean* finds no differences between the literal format of the query produced by the Mediator and the one to be sent to the underlying database. By contrast, in the case of DICOM, a different format is used to instantiate patient identifiers. *OntoQueryClean* translates the original identifier into the proper value for DICOM.

The second example involved a format transformation applied over a different integrated query. In this case, the mapping for this query produced subqueries only for TOP and SIOP. A restriction on the date of a specific treatment on patients was included in the original query. In the global schema, values for dates are separated in three different fields—i.e. day, month and year—but both TOP and SIOP employ a single field with the union of the mentioned fields—with the format DD-MM-YYYY—. We used the format transformation method of *OntoDataClean* to overcome this heterogeneity, successfully producing the expected literals in the restrictions.

It must be also noted that, even though no exact measures were performed, no significant increase in latency of the system was noticed when including the *OntoQueryClean* tool in the Mediator.

4. Conclusions

We have described a method and developed a tool for query heterogeneities preprocessing. The method presented is based on the use of ontologies to classify the most common query identifier heterogeneities, and to automate the query identifier transformation task. A dedicated ontology describing the domain has been described, based on previous works on this task. Initially, three transformations methods are

described—namely: scale, format and synonym transformations. The ontology can be extended easily to support future requirements.

The current implementation of OntoQueryClean tool supports the described methods, which cover the most important problems when dealing with literal heterogeneity in query translation. This tool allows transformation of literals inside SPARQL queries, facilitating the integration of heterogeneous sources. OntoQueryClean has been successfully integrated into the ACGT Semantic Mediator, a mediation system being implemented inside the ACGT project. The tool is composed of two modules: i) the parsing module and ii) the preprocessing module. The former is devoted to parsing SPARQL queries in order to extract the literals that must be transformed, while the latter performs the actual transformations.

Two experiments were presented, based on real queries formulated by clinicians, proving the suitability of our approach. In both cases, the tool successfully generated the expected results, allowing to properly querying the respective data sources.

Future work will involve increasing the features offered by the tool by embracing more transformation methods and by offering support to different query languages. We plan to include machine methods to automate the task of defining the transformation, such as the use of similarity functions. These improvements will lead to a more general purpose tool, able to cope with the requirements imposed by a wider set of domains.

5. Acknowledgements

This work was funded by the European Commission, through the ACGT integrated project (FP6-2005-IST-026996).

References

- [1] D. Perez-Rey, A. Anguita, and J. Crespo: OntoDataClean: Ontology-based Integration and Preprocessing of Distributed Data, *Lecture notes in Computer Science* **4345** (2006), 262-272.
- [2] N. Noy: Order from Chaos, *ACM Queue* vol. 3 **8** (2005). Available at: <http://www.acmqueue.org/modules.php?name=Content&pa=showpage&pid=341&page=1>
- [3] A. Halevy: Why Your Data Won't Mix, *ACM Queue* vol. 3 **8** (2005). Available at: <http://www.acmqueue.org/modules.php?name=Content&pa=showpage&pid=336>
- [4] D. Perez-Rey, V. Maojo, M. Garcia-Remesal, R. Alonso-Calvo, H. Billhardt, F. Martin-Sanchez, and A. Sousa: ONTOFUSION: Ontology-Based Integration of Genomic and Clinical Databases, *Computers in Biology and Medicine* **36** (2006), 712-730.
- [5] J. Köhler, S. Philippi, and M. Lange: SEMEDA: ontology based semantic integration of biological databases, *Bioinformatics* vol 19 **18** (2003), 2420-2427.
- [6] G.R. Librelotto, W. Souza, J.C. Armalo, and P.R. Henriques, Using the Ontology Paradigm to Integrate Information Systems, *Proceedings of the International Conference on Knowledge Engineering and Decision Support*, Porto (2004), 497–504.
- [7] C. Bizer: D2R MAP - A Database to RDF Mapping Language, *Proceedings of the International World Wide Web Conference*, Budapest (2003)
- [8] Sources and Classification of Semantic Heterogeneities. Available at: <http://www.mkbergman.com/?p=232>

Gene Regulation Ontology (GRO): Design Principles and Use Cases

Elena BEISSWANGER^{a1}, Vivian LEE^b, Jung-Jae KIM^b, Dietrich REBHOLZ-SCHUH-MANN^b, Andrea SPLENDIANI^c, Olivier DAMERON^c, Stefan SCHULZ^d, Udo HAHN^a

^a *Jena University Language and Information Engineering (JULIE) Lab, Jena, DE*

^b *European Bioinformatics Institute, Hinxton, Cambridge, UK*

^c *Laboratoire d'Informatique Médicale, Université de Rennes 1, Rennes, FR*

^d *Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Freiburg, DE*

Abstract. The Gene Regulation Ontology (GRO) is designed as a novel approach to model complex events that are part of the gene regulatory processes. We introduce the design requirements for such a conceptual model and discuss terminological resources suitable to base its construction on. The ontology defines gene regulation events in terms of ontological classes and imposes constraints on them by specifying the participants involved. The logical structure of the ontology is intended to meet the needs of advanced information extraction and text mining systems which target the identification of event representations in scientific literature. The GRO has just been submitted to the OBO library and is currently under review. It is available at <http://www.ebi.ac.uk/Rehholz-srv/GRO/GRO.html>

Keywords. Bio-ontologies, Knowledge bases, Terminology-vocabulary

Introduction

In this paper we introduce an ontology, which deals with the conceptual structures underlying gene regulation. It constitutes the ontological backbone of an information extraction and text mining system developed within the framework of the BOOTStrep project,² whose goal is to automatically harvest unstructured information from natural language documents and to structure this information in a biological fact data base.

The processing requirements for information extraction are usually more sophisticated than those for document retrieval or classification tasks. Hence, formal foundations and a high level of expressivity are needed for the representation of domain knowledge in terms of an ontology. These considerations preclude, to a large extent, the direct re-use of existing biomedical vocabularies or terminological resources

¹Corresponding author: Elena Beisswanger, Jena University Language & Information Engineering (JULIE) Lab, Fürstengraben 30, 07743 Jena, Germany; E-mail: elena.beisswanger@uni-jena.de

²<http://www.bootstrep.org>

as available from the *Unified Medical Language System* (UMLS),³ since they are neither formal nor expressive enough. These resources, however, still have to be checked for relevant terminology in the field of gene regulation that subsequently could be enriched with new logical interconnections. In a similar way, the *Open Biomedical Ontologies* (OBO) library⁴ has to be screened, a collection of publicly available biomedical ontologies that has some intersections with the medically focused UMLS.

After a brief introduction into the domain of gene regulation, we will review the resources we used in setting up the *Gene Regulation Ontology* (GRO) (Section 3.1), and then survey the structure of GRO in terms of its basic entity types and relations, closely following the recommendations issued by the OBO Foundry [1].

1. Gene regulation - a brief refresher

Gene regulation, the regulation of gene expression, characterizes the cellular mechanisms that control the amount of gene products of individual genes synthesized at a particular time and under particular extra- and intracellular conditions. Most known mechanisms regulate the expression of protein coding genes. Gene expression falls into two major phases, *viz.* transcription and translation. During transcription, proteins called transcription factors (TF) bind to specific binding sites of a gene. This process starts or stops the transcription of the gene into an RNA, the intermediate product of gene expression, by a polymerase enzyme. In the second part of gene expression, the RNA is translated into a protein. Regulatory processes occur on all of the different steps of gene expression, from transcription to post-translational protein modification. They enable the cell to adapt to different conditions controlling its structure and function. In many cases abnormal regulation of gene expression causes diseases. A prominent example is the induction of cancer in cells, in which abnormal regulation of gene expression plays a crucial role. Our work will particularly focus on the regulation occurring on the transcriptional level of gene expression.

2. The Gene Regulation Ontology (GRO)

GRO has been created as a conceptual model for the domain of gene regulation following best practice design principles as recommended by the OBO Foundry [1]: It is publicly available, uses a *commonly shared syntax*, *viz. the Web Ontology Language (OWL)*,⁵ and has a *clearly specified content*. It covers processes occurring on the intracellular level (such as the binding of TFs to DNA binding sites), and physical entities that are involved in these processes (such as genes and TFs) in terms of ontology classes interlinked by semantic relations. No instances of classes are included in the GRO. As an ontology supporting natural language processing (NLP) applications (see Section 4), the GRO is intended to represent *common knowledge* about the domain and focuses on the relations between types in the domain, rather than representing overly fine-grained classes as can be found in ontologies created for data base annotation purposes, such as the *Gene Ontology* (GO) [2].

³<http://www.nlm.nih.gov/pubs/factsheets/umls.html>.

⁴<http://www.bioontology.org/repositories.html#obo>

⁵<http://www.w3.org/TR/owl-features/>

2.1. Design and Construction of the GRO

The basic structure of the GRO was developed manually based on textbook knowledge and considering the terminology already available within the UMLS resources. To populate GRO and interlink it with existing ontological resources the OBO ontologies were screened by two biologists for entries related to gene regulation. These were subsequently extracted and integrated into the GRO, while keeping the references to the source terminologies. In addition, information was taken from the transcription factor database TransFac [3]. Table 1 lists all selected sources and the kind of information we derived. To complete the conceptual representation of gene regulation, in the next step, 150 Medline⁶ abstracts (selected by a MeSH⁷ query and additional selection criteria) were analyzed with regard to potentially new GRO terms.

The GRO relies on a taxonomic backbone based on *is-a* (subclass) relations between ontology classes. In addition, it provides a set of semantic relation types following and extending the OBO *Relation Ontology* (RO) [4]. The RO has recently become a *de facto* standard for ontology relations in the biomedical domain and using the RO relations will improve the interoperability between different ontologies.

Table 1. Conceptual resources for the population of the GRO

Resource with URL	Relevant Information
Gene Ontology (GO) http://geneontology.org/	molecular functions, biological processes, cellular components
Sequence Ontology (SO) http://sequenceontology.org/	sequence regions and attributes of sequence regions
ChEBI http://www.ebi.ac.uk/chebi/	chemical entities
INOH Molecule Role (IMR) http://www.inoh.org/	transcription factors and their functional domains
NCBI taxonomy, http://130.14.29.110/Taxonomy/	eukaryotes, prokaryotes
TransFac http://www.gene-regulation.com/	transcription factors, domains of transcription factors

The GRO is continuously evolving. The current version (as of February 20, 2008) comprises 433 classes related by 396 semantic relations of 8 different types, exclusive the taxonomic *is-a* relation and inverse / reciprocal relations (see Section 3.2).

2.2. GRO Classes and Relations

Basically the GRO consists of two branches. While the **continuant** branch describes entities ‘which persist through time’, the **occurrent** branch describes process entities, such as **transcription**, **gene expression**, and various regulatory processes. A continuant can either be ‘physical’, i.e., it has a spatial dimension (such as **gene**, **regulatory sequence**, and **protein**), or ‘non-physical’ (such as **protein function**), whereas occurrents are always non-physical.

A feature that distinguishes the GRO from most other biomedical ontologies is that its classes are highly interlinked by various manually encoded relations (reciprocal relations are given in brackets, hereafter). The relation *part-of* (*has-part*) is used to

⁶<http://www.nlm.nih.gov/pubs/factsheets/medline.html>

⁷<http://www.nlm.nih.gov/pubs/factsheets/mesh.html>

relate spatial or temporal parts to the whole, such as the class **protein domain** to the class **protein** or the class **transcription initiation** to the class **transcription**. Wholes and their parts must belong to the same ontological category, i.e., a continuant can only have continuants as parts and an occurrent can only have occurrents as parts. The relation *from-species* relates species-specific classes to the species they refer to, such as **bacterial RNA polymerase** refers to **bacterium**. Also, continuants and occurrents are related to processes in which they are involved by the relation *participates-in* (*has-participant*), or by one of the two sub-relations *agent-of* (*has-agent*) or *patient-of* (*has-patient*), respectively. Consider, e.g. the event **regulation of transcription** – in the GRO it is defined as a subclass of a **regulation of gene expression** with the following restrictions: (i) **regulation of transcription** *has-agent* **transcription regulator**, and (ii) **regulation of transcription** *has-patient* **transcription**. Finally, the relation *encodes* (*encoded-in*) relates genes to proteins, *function-of* (*has-function*) links functions to their bearers, *has-quality* specifies qualities inherent in particular entities, and *results-in* (*results-from*) identifies the outcome of a process. To further exploit the potential of formal ontologies, domain and range constraints, as well as the algebraic properties (such as transitivity, reflexivity) of the relations need to be defined in the GRO in a pending revision step.

2.3. Modeling Decisions and Implementation

Exactly as the GO and the *Sequence Ontology* (SO), which are both members of the OBO, the GRO is designed in a species-independent manner. However, since major differences between eukaryotic and prokaryotic gene regulation exist, they have to be reflected by the ontological representation. To allow for species-specific classes in GRO we introduced the relation *from-species*. An advantage of this approach is that it does not require a separate ontology for every particular species. In contrast to the GO, which is basically used by human curators, the GRO is designed for NLP applications (Section 4), which require, e.g., to distinguish between functions and function bearers. Thus, the GRO holds both, functions as described in the GO molecular function branch, and classes representing function bearers, linked by *has-function* relations.

The GRO is implemented in OWL-DL, the description logic variant of OWL. OWL offers mechanisms to define classes and to relate them according to their properties. The formal definition of a typical GRO class holds a Uniform Resource Identifier (URI), a human readable class name provided by the predefined OWL property `rdfs:label`, a textual definition explaining the meaning of the class, at least one *is-a* relation, and references to similar classes in external resources such as the OBO ontologies. In addition, OWL class restrictions which involve further relations (e.g., *has-agent*, *has-patient*) are used to logically constrain the meaning of a class. OWL supports negation as well as existential and universal constraints.

Classes which do not share any instances are disjoint. To enable more restrictive consistency checks in GRO, we exploited OWL's support to explicitly express class disjointness and added corresponding statements to the GRO wherever this was necessary. We marked, e.g., the classes **DNA** and **RNA** as disjoint because an existing nucleotide sequence, by nature, cannot belong to both classes, whereas the classes **cell** and **organism** are not disjoint because they share instances, such as bacterial cells. There are some limitations of OWL-DL that we had to deal with when implementing the ontology. For example, n-ary relations are not supported, but usually they can be expressed by a set of binary relations. Further, neither statements such as “some

process X *has-agent* some protein Y”, nor uncertainty, nor exceptions can be expressed in OWL. In the BOOTStrep project, we decided to move this kind of information into a separate biological fact data base that is linked to the GRO.

3. Applications of the Gene Regulation Ontology

Bio-ontologies are designed to meet the needs of predefined use cases (e.g., annotation of genetic sequences). GRO is intended to be used as the conceptual backbone for automatic document analysis, information extraction and text mining tasks, in particular. This imposes specific requirements on GRO which are different from those for document retrieval or classification tasks (typical applications of the UMLS). They also differ from those for functional annotations in bio databases, the common application area for many of the OBO ontologies.

3.1. Vocabulary for the Semantic Annotation of Scientific Documents

In the BOOTStrep project, a proper subset of the GRO terms is used as annotation vocabulary for the semantic annotation of biological documents. Corpora annotated by such semantic meta data are a prerequisite for supervised machine learning algorithms. In BOOTStrep, semantic annotations are carried out on several levels. The bottom level is the annotation of entities, i.e., the assignment of labels from a controlled vocabulary to the participants of gene regulation, such as TFs and genes. The vocabulary is taken from the **continuant** branch of the GRO. The next level of annotation relates to regulatory processes (event annotation), a much more complex task, which requires entity annotations as a prerequisite. Terms from the **occurrent** branch of the GRO are used as vocabulary and participation relations specified for these terms are exploited to constrain the assignment of semantic roles. There is some evidence here that annotation tasks like the one outlined above might profit from an ontology-based annotation vocabulary [5].

3.2. Basis for SWRL Rules to Derive Biological Knowledge

The reasoning facilities that come with the OWL language family are needed for additional processing tasks like the analysis of complex representations of events that require the identification of participants involved and their relations. We are currently investigating into a system that extracts mentions of events involving gene regulations from the scientific literature by using the GRO. The system first recognizes entities such as gene and protein names and labels them with classes in the **continuant** branch of the GRO, particularly with the two classes **gene** and **transcription regulator**. The entity recognition utilizes UniProtKB⁸ and RegulonDB⁹ as sources for names of the two classes. The system then identifies instances of the classes in the **occurrent** branch by matching linguistic patterns of keywords of the classes (e.g. ‘regulate’ for **regulatory process**). It finally deduces complex information by employing rules written in SWRL¹⁰ thus capturing sophisticated forms of biological knowledge. For

⁸<http://www.ebi.ac.uk/uniprot/>

⁹<http://regulondb.ccg.unam.mx/>

¹⁰<http://www.w3.org/Submission/SWRL/>

example, the system may identify two instances of the classes **regulation of gene expression** and **binding of transcription factor to DNA**, which share instances of **transcription regulator** and **gene** as their agent and patient, respectively. It will then make the rule-based deduction that the instance of **regulation of gene expression** also belongs to the class **regulation of transcription** which is a subclass of **regulation of gene expression**. We are developing such rules, which should be OWL-DL safe [6], based on the classes and relations provided by the ontology. The GRO is necessary to implement these rules, since such inferences cannot be achieved without a logically sound ontology

4. Conclusions and Outlook

We have introduced the *Gene Regulation Ontology* GRO, a conceptual model for the regulation of gene expression. It covers both, gene regulatory processes occurring on the intracellular level and molecular entities participating in these processes. The GRO was created within the BOOTStrep project aiming at the integration of biomedical knowledge at a very large scale. In particular the GRO serves as controlled vocabulary for semantic annotation of biomedical texts which forms the basis for knowledge-intensive NLP tasks, such as rule-based high quality information extraction.

In the next period of work the participation relations provided by the ontology will be refined to enable an even more fine-grained process modeling. Further on, the expressivity of the ontology will be increased by defining domain and range constraints of the already existing relations, as well as their algebraic properties (such as transitivity, reflexivity). We expect that besides using the GRO in support of NLP tasks it might become relevant for ongoing biological research, too, raising interest and demands for its use in the classification of genes and transcription factors according to the processes they are involved in. In this way, the GRO would complement the Gene Ontology by formally describing what a particular GO term ‘means’ in terms of specifying relations between participants and the processes in which they are involved.

Acknowledgments

The work presented here is part of the BOOTStrep project, a Specific Targeted Research Project (STREP) of the 6th Framework Program of the European Union (FP6 - 028099).

References

- [1] B. Smith, M. Ashburner, C. Rosse, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25(11):1251–1255, 2007.
- [2] The Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, 32(1):258–261, 2004.
- [3] P. Stegmaier, A. E. Kel, E. Wingender. Systematic DNA-Binding Domain Classification of Transcription Factors. *Genome Informatics*, 15(2): 276–286, 2004.
- [4] B. Smith, W. Ceusters, B. Klagges, et al. Relations in biomedical ontologies. *Genome Biology*, 6(5):R46 (1:15), 2005.
- [5] A. Kawazoe, L. Jin, M. Shigematsu, et al. The development of a schema for the annotation of terms in the BioCaster disease detec./trac. system. Proceedings of the KR-MED 2006 Workshop, Baltimore, MD, USA, Nov. 2006, pp.77-85
- [6] B. Motik, U. Sattler, R. Studer. Query answering for OWL-DL with rules. Proceedings of the 3rd International Semantic Web Conference (ISWC 2004), Hiroshima, Japan, November 2004, pp. 549-563.

Differences in Doctors' and Nurses' Assessments of Hospital Culture and their Views about Computerised Order Entry Systems

Joanne CALLEN^{a,1}, Jeffrey BRAITHWAITE^b, Johanna WESTBROOK^a,

^a *Health Informatics Discipline, The University of Sydney, Australia*

^b *Centre for Clinical Governance Research in Health, the University of New South Wales, Australia*

Abstract. The organisational culture of a health facility has been identified as a significant factor for successful implementation of clinical information systems. There have been no reported studies exploring the link between sub-cultures and the use of information systems. This study utilises cross sectional surveys to measure doctors' and nurses' perceptions of organisational culture and relate this to their use of a hospital-wide mandatory computerised pathology order entry (CPOE) system. Data were collected by administering an organisational culture survey (Organisational Culture Inventory, OCI) along with a user-satisfaction survey to a population of 103 doctors and nurses from two clinical units in an Australian metropolitan teaching hospital. We identified subcultures based on professional divisions where doctors perceived an aggressive/defensive culture (mean percentile score = 43.8) whereas nurses perceived a constructive culture (mean percentile score = 61.5). There were significant differences between doctors and nurses on three of the attitude variables with nurses expressing more positive views towards CPOE than doctors. The manifestation of subcultures within hospitals and the impact this has on attitudes towards clinical information systems should be recognized and addressed when planning for system implementation.

Keywords. Implementation - deployment, Diffusion, Managing Change, Organisation,

Introduction

Information and communication technologies in the health industry have been slow to diffuse, particularly in relation to patient care clinical information systems [1]. The organisational culture of a health facility has been identified as a significant factor for successful implementation of such information systems [2-4]. Organisational cultures

¹ Joanne Callen, PO Box 170 Lidcombe NSW 1825 Australia. j.callen@usyd.edu.au
www.fhs.usyd.edu.au/hireu

in health are complex and pluralist with typically multiple cultural features operating simultaneously. Sub-cultures inherent in health care organizations [5,6] may be based on occupational, professional, gender, race or functional distinctions. They can reflect or differentiate from enterprise-wide organisational culture and can draw their values and beliefs from outside ties, for example medical subspecialty professional groups and colleges [7].

As the value of clinical information systems is becoming increasingly evident, what is lacking is a study which measures hospital sub-cultures and links this to the mandatory use of a clinical information system. This paper seeks to provide such a study. We investigated the following question: Is there a relationship between hospital sub-cultures and attitudes to, and satisfaction with, clinical information technology? There have been no published studies we could find which have explored this relationship.

1. Methods

1.1. Research Settings

The study was conducted in two clinical units, an Emergency Department and Haematology/Oncology ward, in a 405 bed Australian metropolitan public teaching hospital. The choice of hospital for the study was purposive, based on its long-term use of a CPOE system. This was a hospital-wide, mandatory system which allowed doctors and nurses to order and view clinical laboratory and radiology tests electronically for all patients. Doctors and nurses were the primary users of CPOE, with each having different levels of use.

1.2. Study design, population and data collection

A cross sectional survey design was employed using two instruments: an organisational culture inventory; and a user satisfaction survey. The population (n=103) comprised all doctors (n=42) and nurses (n=61) in the two clinical units of the hospital.

1.3. Study instruments

1.3.1. The Organisational Culture Inventory

The Organisational Culture Inventory (OCI) [8] provides a point-in-time picture of clinicians' perceptions of the culture of their hospital in terms three general clusters which distinguish between: **constructive** (members are encouraged to work cooperatively and to their full potential resulting in high levels of motivation and teamwork); **passive/defensive** (members are expected to please those in positions of authority and wait for others to act first); and **aggressive/defensive** (members are expected to oppose new ideas, compete and appear competent and independent) cultures [9]. Within each of the three clusters there are four behavioural norms which constitute the cultural style. For example within the **constructive culture** the four norms are: **achievement** (expected to set challenging but realistic goals); **self-actualising** (be concerned about their own growth); **humanistic/encouraging** (be

supportive and resolve conflicts constructively); and **affiliative** (expected to cooperate with others).

1.3.2. User satisfaction survey

A user satisfaction questionnaire based on previous point of care evaluation questionnaires [10] was developed and trialled with the final survey consisting of twenty two closed -ended questions relating to the impact of CPOE on work practices, patient care, and doctors' and nurses' satisfaction with, and attitudes to, the system.

1.4. Data analysis

Percentile scores were calculated for each of the 12 behavioural norms in the OCI and significance was tested using χ^2 . Scores below the 50th percentile are low and reflect weak expectations of the behaviour in question. Scores above the 50th percentile reflect stronger expectations and scores that fall close to the 50th percentile reflect moderate expectations for the behaviour in question. The means of the percentile scores of the four culture styles within each culture cluster were calculated to ascertain the culture of each professional group.

2. Results

2.1. Population and respondents

From the population of 103 clinicians in the two clinical units, 75 (30 doctors and 45 nurses) completed the OCI (response rate = 73%) and 96 (36 doctors and 60 nurses) completed the user satisfaction survey (response rate = 93%).

2.2. Comparisons between doctors' and nurses' perception of culture

There were differences between the perceptions of doctors and nurses where doctors perceived an *aggressive-defensive culture* (mean percentile score = 43.8) in contrast to nurses who perceived a *constructive culture* (mean percentile score = 61.5). There was a significant difference between the doctors' (percentile score = 35) and nurses' (percentile score = 80) perceptions on the *humanistic/encouraging style* ($p < 0.01$). Doctors were also high on the *perfectionistic style* (percentile score = 61).

2.3. Comparisons between doctors' and nurses' attitudes to, and satisfaction with, computerised test management systems

There were significant differences between responses of doctors and nurses on three attitude variables (Table 1). Nurses were significantly more likely to agree that computerised test management systems helped in deciding which tests are appropriate to order and made their work more interesting. Doctors on the other hand were more likely to agree that CPOE resulted in over-ordering of tests.

Table 1. Attitudes to the computerised test management system (n=96)

Attitudes ^a	Doctors n=36			Nurses n=60 ^b			p-value
	Agree % (n)	Neutral % (n)	Disagree % (n)	Agree % (n)	Neutral % (n)	Disagree % (n)	
Using computerised test management systems:							
Results in cookbook medicine	28 (10)	11 (4)	61 (22)	25 (15)	28 (17)	47 (28)	0.134
Depersonalises medicine	11 (4)	19 (7)	69 (25)	10 (6)	12 (7)	78 (47)	0.550
Helps in deciding which tests to order	25 (9)	19 (7)	56 (20)	71 (42)	19 (11)	10 (6)	0.000 ^c
Improves the practice of medicine	61 (22)	31 (11)	8 (3)	73 (44)	17 (10)	10 (6)	0.281
Results in over ordering of tests	58 (21)	14 (5)	28 (10)	30 (18)	38 (23)	32 (19)	0.010 ^c
Alienates doctors from patients	3 (1)	8 (3)	89 (32)	6 (4)	29 (11)	65 (45)	0.254
Makes my work more interesting	33 (12)	50 (18)	17 (6)	63 (31)	29 (14)	8 (4)	0.024 ^c

Notes

1. ^aattitude variables were measured on a five point Likert scale collapsed to a three point Likert scale with 'agree' including 'strongly agree' and 'disagree' including 'strongly disagree'

2. ^bn=60 with 1 missing response from 'helps with deciding which tests to order'

3. ^csignificant at p<0.05

Doctors and nurses were similarly satisfied on all the items measuring satisfaction with ordering tests using the computerised test management system. The majority of doctors and nurses thought that computerised ordering: was reliable (83% and 91%); improves productivity (72% and 66%); was not time consuming (61% and 63%); and did not have a negative impact on patient care (83% and 81%).

3. Discussion and conclusion

Our study showed that organisational cultures are not uniform within large teaching hospitals. This finding supports previous studies on cultures within healthcare organisations as being diverse with numerous emergent sub-cultures, often manifesting along professional lines [11,12]. These results also draw attention to comparative attitudes about the acquisition and use of clinical information systems which have implications for preparing organisations for innovations. They highlight the importance of identifying, explaining and measuring sub-cultures, particularly amongst

professional groups, before implementation to develop tailored strategies to facilitate acceptance and use.

How do organisational sub-cultures as perceived by doctors and nurses in this study relate to their attitudes toward, and satisfaction with, the computerised test management system? The relatively *constructive culture* of nurses is one where they broadly support the uptake and implementation of a new clinical information system. Nurses did in fact have more positive views. The *aggressive/defensive culture* perceived by doctors would appear to discourage the uptake of new innovations as they would by definition tend to oppose new ideas either openly or indirectly, or both. The high *perfectionistic* styles of doctors could also be counterproductive to the implementation of new clinical information systems. A culture of blame and the need to be perfect is not conducive to implementing change.

These findings have implications for those interested in facilitating the acceptance and implementation of clinical information systems. The results indicate the importance of examining sub-cultures prior to implementation to enable cultural characteristics and differences between professional groups to be taken into account. A cross-sectional measure of organisational culture within professional groups provides a point-in-time picture which can be used to prepare for an information technology implementation by highlighting areas where cultural change might be enacted. Certain behaviours can then be encouraged or discouraged. Systems and processes and practices which promote a *constructive culture* can be instituted at the individual/job level (such as goal setting, job design and motivational processes), the manager/unit level (such as performance appraisal) and the organisational level (such as communication and employee involvement). The acknowledgement by governments, policymakers and managers of the importance of organisational sub-cultures for clinical information system adoption will enable more resources and specific policies to be directed towards successful implementation. Cultural readiness for an innovation should be viewed as a key requirement for hospitals when preparing for system implementations with resources and efforts directed towards developing constructive cultures which support innovations. The results from the exploration of organisational sub-cultures at this hospital exposed the relationships between sub-cultures and how doctors' and nurses' perceived and supported innovations. The complexity and characteristics of sub-cultures within large organisations should be taken into account given the impact of culture on the uptake of new technologies.

3.1. Limitations of research methods and procedures

A limitation of this examination of organisational culture at the hospital was that only quantitative data sets were used. These data could be complemented by interviews and observations which might explore the differences between espoused and enacted values and also gauge the strength of attitudes. Another limitation is that the sample included doctors and nurses from two units of one hospital. The results may not be strictly generalisable to the population of doctors and nurses outside the study areas.

Acknowledgements

We thank Nerida Creswick BAppSc (HIM) Hons for her assistance in data collection. JW is supported by an Australian Government National Health and Medical Research Council (NH&MRC) fellowship.

References

- [1] J.S Ash, D.W. Bates, Factors and forces affecting EHR system adoption: Report of a 2004 ACMI discussion, *J Am Med Inform Assoc.* 12 (1) (2005) 18-20.
- [2] A. S. Gosling, J.I. Westbrook, E.W. Coiera, Variation in the use of online clinical evidence: a qualitative analysis, *Int J Med Inf.* 69 (2003)1-16.
- [3] J. Aarts, H. Doorewaard, M. Berg, Understanding implementation: the case of a computerized physician order entry system in a large Dutch university medical center, *J Am Med Inform Assoc.* 11 (3) (2004) 207-16.
- [4] J. L. Callen, J. Braithwaite, J. I. Westbrook, Cultures in hospitals and their influence on attitudes to, and satisfaction with, the use of clinical information systems, *Soc. Sci & Med.* 65 (2007) 635-639.
- [5] T. Scott, R. Mannion, H. Davies, M. Marshall, Implementing culture change in health care: Theory and practice, *Int J Qual Health Care* 15 (2) (2003b)111-118.
- [6] J. Braithwaite, Analysing structural and cultural change in acute settings using a Giddens-Weick paradigmatic approach, *Health Care Anal.* 14 (2006a) 91- 102.
- [7] H. T. O. Davies, Understanding organizational culture in reforming the National Health Service, *J R Soc Med.* 95 (3) (2002)140-142.
- [8] R. A. Cooke, J. C. Lafferty, *Organizational Culture Inventory*, Human Synergistics, Plymouth, MI, 1989.
- [9] J. L. Szumal, *Organisational culture inventory. Interpretation and development guide*, Human Synergistics, Australia, 1998.
- [10] J. G. Anderson, C. E. Aydin, S. J. Jay (Eds.), *Evaluating health care information systems: Methods and applications*, Sage Publications, London, 1994.
- [11] T. Scott, R. Mannion, M. Marshall, H. Davies, Does organisational culture influence health care performance? A review of the evidence, *J Health Serv Res Policy*, 8 (2003) 105-117.
- [12] J. Braithwaite, M. Westbrook, Rethinking clinical organisational structures: an attitude survey of doctors, nurses and allied health staff in clinical directorates, *Journal of Health Services Research & Policy*, 10 (1) (2005) 10-17.

Novelty Detection using One-class Parzen Density Estimator. An Application to Surveillance of Nosocomial Infections

Gilles COHEN^{a,1}, Hugo SAX^b, Antoine GEISSBUHLER^a

^a*Medical Informatics Service, University Hospital of Geneva, 1211 Geneva, Switzerland*

^b*Department of Internal Service, University Hospital of Geneva, 1211 Geneva, Switzerland*

Abstract. Nosocomial infections (NIs) - those acquired in health care settings - represent one of the major causes of increased mortality in hospitalized patients. As they are a real problem for both patients and health authorities, the development of an effective surveillance system to monitor and detect them is of paramount importance. This paper presents a retrospective analysis of a prevalence survey of NIs done in the Geneva University Hospital. The objective is to identify patients with one or more NIs based on clinical and other data collected during the survey. In this classification task, the main difficulty lies in the significant imbalance between positive and negative cases. To overcome this problem, we investigate one-class Parzen density estimator which can be trained to differentiate two classes taking examples from a single class. The results obtained are encouraging: whereas standard 2-class SVMs scored a baseline sensitivity of 50.6% on this problem, the one-class approach increased sensitivity to as much as 88.6%. These results suggest that one-class Parzen density estimator can provide an effective and efficient way of overcoming data imbalance in classification problems.

Keywords. Pattern Recognition, One-Class Classification, Parzen density estimator, Data Imbalance, Nosocomial Infections, Infection control.

Introduction

A major and constant concern in Health Care Institutions is Infection control, particularly of nosocomial² origin, which directly engage the hospital responsibility.

¹ Corresponding Author: Gilles COHEN, Service d'Informatique Médicale, Hôpitaux Universitaires de Genève, Rue Micheli-du-Crest, 24, Genève, Switzerland, E-mail: gilles.cohen@sim.hcuge.ch

² A nosocomial infection is a disease that presents itself in hospitalized patients in whom the infection was not present at the time of the admission.

This leads to focus increasingly on surveillance to detect and monitor infections, nosocomial or not. Data are collected to assess the magnitude of the problem, detect outbreaks, identify risk factors, target control measures on high-risk patients or wards, and evaluate prevention programs. Finally, the surveillance aims to decrease infection risk and consequently improve patients' safety.

Two methods are generally used for surveillance: (1) trans-sectional assessment (i.e. prevalence studies), which gives estimates on a large population at relatively low cost; or (2) prospective, ongoing surveillance (incidence studies). The latter is the method of choice, however, this method is arduous, unfeasible at a hospital level, and currently recommended only for high-risk, i.e., critically ill patients. Prevalence surveys, which constitute an alternative and more realistic approach, are being recognized as a valid surveillance strategy and are increasingly used. Their major limitations are their retrospective nature, the dependency on readily available data, a prevalence bias, the inability to detect outbreak (depending on the frequency of the surveys), and the limited capacity to identify risk factors. However, the data they provide are sufficient to measure the magnitude of the problem, evaluate a prevention program, and help to allocate resources. They give a snapshot of clinically active NIs during a given index day and provide information about the frequency and characteristics of these infections. The efficacy of infection control policies can be easily measured by repeated prevalence surveys [1].

1. Data collection and preparation

The University Hospital of Geneva (HUG) has been performing yearly prevalence studies since 1994 [2]. Their methodology is as follows: the investigators visit every ward of the HUG over a period of approximately three weeks. All patients hospitalized for 48 hours or more at the time of the study are included. Medical records, X-ray and microbiology reports are reviewed, and additional information is eventually obtained by interviewing nurses or physicians in charge. Each nosocomial infection is recorded according to modified Centres for Disease Control criteria. The survey includes only infections still active at any point during the six days preceding the visit. Collected variables include demographic characteristics, admission date, admission diagnosis, comorbidities, McCabe score, type of admission, provenance, hospitalization ward, functional status, previous surgery, previous intensive care unit stay, exposure to antibiotics, antacid and immunosuppressive drugs and invasive devices, laboratory values, temperature, date and site of infection, fulfilled criteria for infection. Although less time-consuming than prospective surveillance, a prevalence survey nevertheless requires considerable resources, i.e., approximately 800 hours for data collection and 100 hours for entering data in an electronic data base. What is particularly time-consuming is the careful examination of each available information for all patients, in order to detect those who might be infected. This pilot study was aimed at applying machine learning techniques to data collected in the 2002 prevalence survey in order to detect nosocomial infections using the factors described above.

The dataset consisted of 688 patient records and 83 variables. With the help of hospital experts on nosocomial infections, we filtered out spurious records as well as irrelevant and redundant variables, reducing the data to 683 cases and 49 variables. In addition, several variables had missing values, due mainly to erroneous or missing measurements. These values were assumed to be missing at random, as domain experts

did not detect any clear correlation between the fact that they were missing and the data (whether values of the incomplete variables themselves or of others). We replaced these missing values with the class-conditional mean for continuous variables and the class-conditional mode for nominal ones.

2. The imbalanced data problem

The major difficulty inherent in the data (as in many medical diagnostic applications) is the highly skewed class distribution. Out of 683 patients, only 75 (11% of the total) were infected and 608 were not. The problem of imbalanced datasets is particularly crucial in applications where the goal is to maximize recognition of the minority class³. The issue of class imbalance has been actively investigated and remains largely open; it is handled in a number of ways [3], including: oversampling the minority class, building cost-sensitive classifiers [4] that assign higher cost to misclassifications of the minority class, stratified sampling on the training instances to balance the class distribution [5] and rule-based methods that attempt to learn high confidence rules for the minority class [6]. In this paper we investigate another way of biasing the inductive process to boost sensitivity (i.e., capacity to recognize positives) based on one-class Parzen density estimator. Experiments conducted to assess this approach are described in Section 5 and results are discussed in Section 6.

3. Classification algorithm

3.1. One-class classification

In typical classification problems a discriminative hypothesis h is found based on training cases from all c (≥ 2) classes, so that it can classify each new unseen case into one of c classes with the lower possible generalisation error.

While many pattern recognition problems fall into this category, some other problems are best formulated differently as *one-class* or *novelty detection* problems. The one-class classification task can be expressed as the ability to distinguish between new cases similar to members of the training set and all other cases that can occur. In a probabilistic sense, one-class classification is equivalent to deciding whether a previously unseen test case has been produced by the underlying distribution of the training set of normal cases. While it seems to be similar to conventional classification problems (i.e. two-class), one-class classification differs in the way the classifier is trained. It is trained only by cases from the majority class, and never sees those from the minority class. It must estimate the boundary that separates those two classes based only on data which lies on one side of it. The one-class approach is particularly attractive in situations where cases from one class are expensive or difficult to obtain for model construction. In a one-class setting, novel or abnormal cases can be detected by constructing a real-valued density estimation function. The most straightforward method is to estimate the density of the training data and to set a threshold on this

³ For convenience we identify positive cases with the minority and negative cases the majority class.

density. In this study such an approach is used : a one-class Parzen density estimator. We briefly describe it in the following section.

3.2. One-class Parzen Density Estimator

The Parzen Window [7,8,9] is a simple kernel based estimator for estimating density function. The main idea of this non-parametric method is to place a Gaussian kernel for each pattern within the pattern vector determining the position of the center of the kernel. The Gaussian kernels have a smoothing parameter that control the smoothness of the kernels to give appropriate classification performance. The smoothing parameter is normally global to the entire set of Gaussian kernels. Let $p(x)$ be the density function to be approximated. Consider a training set X of n i.i.d. examples drawn according to $p(x)$, the Parzen window estimate of $p(x)$ based on the X is $\hat{p}(x) = \frac{1}{n} \sum_{i=1}^n \kappa(x - x_i)$

where $\kappa(\cdot)$ is a kernel function with localized support and its exact form depends on n . The most commonly used is the Gaussian kernel :

$$\hat{p}(x) = \frac{1}{n(2\pi)^{d/2} \sigma^d} \sum_{i=1}^n \exp \left\{ -\frac{\|x - x_i\|^2}{2\sigma^2} \right\} \text{ here } x_i \text{ is an example in the training set, } \sigma$$

is the smoothing function or bandwidth and d is the dimensionality of the feature space. The density estimator $\hat{p}(x)$ obtained from the training set give us a quantitative measure of the degree of novelty for each new example. This is used to reject examples where the estimate $\hat{p}(x) < \rho$ for some threshold ρ , effectively generating a new class of “novel” data. Thus any point where the likelihood $\hat{p}(x)$ is below some threshold is considered to be novel. This approach was used by Bishop [10] who gives the following justification : Denote C_1 as normality and C_2 as novelty. The corresponding prior probabilities are $P(C_1)$ and $P(C_2)$ and the probability density function are $p(x|C_1)$ and $p(x|C_2)$. According to Bayes, $x \in C_1$ if $P(C_1|x) = p(x|C_1) P(C_1) > P(C_2|x) = p(x|C_2) P(C_2)$. Deciding whether a new example x is abnormal or novel depends on the comparison between $p(x|C_1)$ and $p(x|C_2)P(C_2)/P(C_1)$, where the latter is equivalent to a threshold ρ on $\hat{p}(x)$.

4. Experimental Setup

4.1. Evaluation strategy

The experimental goal was to assess the ability of one-class Parzen density estimator to cope with imbalanced datasets. To train one-class Parzen density estimator we experimented with different values for the parameter ρ for a fixed σ obtained by maximizing the “pseudo-likelihood” computed using cross validation (see Eq. 1).

$$\sigma = \operatorname{argmax}_{\sigma} \left\{ \frac{1}{n} \sum_{i=1}^n \log(f(x_i)) \right\} \text{ where } f(x_j) = \frac{1}{(n-1)\sigma} \sum_{i=1, i \neq j}^n \kappa \left(\frac{x_j - x_i}{\sigma} \right) \quad (1)$$

For both approaches generalization error was estimated using 5-fold cross-validation. The complete dataset was randomly partitioned into five subsets. On each iteration, one subset (comprising 20% of the data samples) was held out as a test set and the remaining four (80% of the data) were concatenated into a training set. Note that in this approach, the training sets consisted only of non infected patients whereas the test sets contained both infected and non infected patients according to the original class distribution. Error rates estimated on the test sets were then averaged over the five iterations. Overall performance was quantified using the metrics discussed in the following section.

4.2. Performance Metrics

To discuss alternative performance criteria we adopt the standard definitions used in binary classification. TP and TN stand for the number of true positives and true negatives respectively, i.e., positive/negative cases recognized as such by the classifier. FP and FN represent respectively the number of misclassified positive and negative cases. In two-class problems, the accuracy rate on the positives, called sensitivity, is defined as $TP/(TP+FN)$, whereas the accuracy rate on the negative class, also known as specificity, is $TN/(TN+FP)$. Classification accuracy is simply $(TP + TN)/N$, where $N=TP+TN+FP+FN$ is the total number of cases.

5. Results

Table 1. Performance of one-class Parzen density estimator for different parameter settings using an RBF Gaussian kernel. ρ_{perc} is the percentile corresponding to the threshold ρ It means that the lowest ρ_{perc} of the training data is rejected by applying the corresponding threshold.

Parameters		Accuracy	Sensitivity	Specificity
σ	ρ_{perc}	%	%	%
0.14	5	65.7	12.9	98.3
	10	69	23.6	97.1
	20	72	39.6	92
	40	72,9	68.3	75.8
	50	70.6	88.6	59.5

Table 2. Best performance of SVMs with symmetrical and asymmetrical margin.

SVM Classifier	Accuracy	Sensitivity	Specificity
Sym. Margin	89.6%	50.6%	94.4%
Asym. Margin	74.4%	92%	72.2%

Tables 1 summarize performance results for the one-class Parzen density estimator. They show the best results obtained by training classifiers using different parameter configurations on non infected cases only.

Clearly, for the one-class Parzen density estimator, highest sensitivity is reached when ρ corresponds to a rate of training data rejection of 50%; the price to pay for such

a result is that many non infected cases are equally labelled abnormal, thus yielding low specificity. One-class approach lead to significant improvements in sensitivity over classical symmetrical SVMs. Moreover a major limitation of one-class Parzen density estimator is its relatively high computational demand during the testing phase.

In a previous study on the same nosocomial dataset [11], we investigated a support vector algorithm in which asymmetrical margins are tuned to improve recognition of rare positive cases. Table 2 shows the best performance measures obtained in these previous experiments.

6. Conclusion

We analyzed the results of a prevalence study of nosocomial infections in order to detect infected patients. The major hurdle, typical in medical diagnosis, is the problem of rare positives. To address this problem we investigated the applicability of a one-class algorithm proposed by [10]. Experimental results reported in this paper are encouraging. From the point of view of sensitivity, one-class Parzen density estimator attain the highest level (88.6 %) observed by the authors throughout a series of studies on the problem. However, the price paid in terms of loss in specificity is quite exorbitant, and domain experts must decide if the high recognition rate is worth the cost of treating false positive cases. From this point of view, asymmetrical-margin SVMs might prove preferable in that they maintain a more reasonable sensitivity-specificity trade-off. Overall we feel that one-class Parzen density estimator are a promising approach to the detection of nosocomial infections and can become a reliable component of an infection control system.

References

- [1] G. G. French, A. F. Cheng, S. L. Wong, and S. Donnan. Repeated prevalence surveys for monitoring effectiveness of hospital infection control. *Lancet*, 2:1021–23, 1983.
- [2] S. Harbarth, C. Ruef, P. Francioli, A. Widmer, D. Pittet, and S.-N. Network. Nosocomial infections in Swiss university hospitals: a multi-centre survey and review of the published experience. *Schweiz Med Wochenschr*, 129:1521–28, 1999.
- [3] N. Japkowicz. The class imbalance problem: A systematic study. *Intelligent Data Analysis Journal*, 6(5), 2002.
- [4] P. Domingos. A general method for making classifiers costsensitive. In *Proc. 5th International Conference on Knowledge Discovery and Data Mining*, pages 155–164, 1999.
- [5] M. Kubat and S. Matwin. Addressing the curse of imbalanced data sets: One-sided sampling. In *Procs of the Fourteenth International Conference on Machine Learning*, pages 179–186, 1997.
- [6] K. Ali, S. Manganaris, and R. Srikant. Partial classification using association rules. In *Proc. 3rd International Conference on Knowledge Discovery in Databases and Data Mining*, 1997.
- [7] B. Silverman. *Density Estimation*. Chapman and Hall, 1986.
- [8] E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.
- [9] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley, 2000.
- [10] C. M. Bishop. Novelty detection and neural network validation. *Vision, Image and Signal Processing, IEE Proceedings-*, 141(4):217–222, 1994.
- [11] G. Cohen, M. Hilario, H. Sax, and S. Hugonnet. Asymmetrical margin approach to surveillance of nosocomial infections using support vector classification. In *Intelligent Data Analysis in Medicine and Pharmacology*, 2003.

Interface Terminologies: Bridging the Gap between Theory and Reality for Africa

Andrew S. KANTER, MD MPH^{a,b}; Amy Y. WANG, MD^b; Fred E. MASARIE, MD^b;
Frank NAEYMI-RAD, PhD MBA^b; Charles SAFRAN, MD PhD^b

^a *International Partnership for Health, NFP. Chicago, USA*

^b *Intelligent Medical Objects, Inc., Chicago, USA*

Abstract. In the United States and Europe, electronic health records (EHRs) allow information technology and decision-support to facilitate the activities of clinicians and are considered an important component of health care improvement. However, actual adoption of EHRs by physicians has been slow and the use of decision support has been minimal. Part of the difficulty lies in the challenges that users face in capturing structured clinical information. Reference and administrative terminologies have been in use for many years and provide a critical infrastructure to support reimbursement, decision-support and data analysis. The problem is that physicians do not think and work using reference terminologies. Interface terminologies bridge the gap between information that is in the physician's mind and information that can be interpreted by computer applications. The maps from interface terminologies to appropriate reference terminologies enable advanced functionality in clinical information systems. The conflict between the need for timely adoption of health information technology and the need for standardization is also relevant to the problems faced by health information technology in Africa. The problem of clinicians having to communicate and/or record information in a format that is acceptable to *someone else, somewhere else*, leaves the true benefits of these systems beyond the reach of most in Africa. There is a growing effort in the United States to produce clinically-relevant interface terminologies mapped to standards. These interface terminologies can be expanded to incorporate the languages and clinical requirements of clinicians in Africa. The adoption of interface terminologies will help bring the value of standard terminology and the resulting benefits of decision-support, data analysis and information retrieval to parts of the world where they are needed most.

Keywords: Terminology-Vocabulary, Standards, EPR-CPR-EMR, Public Health Informatics, Africa

Introduction

Electronic health records (EHRs) are being implemented worldwide with the promise that they will not only dramatically improve the health care of individuals but also will expand knowledge about diseases and treatments, strengthen understanding about the effectiveness and efficiency of health care systems, and support public health. For health information systems to realize these benefits, however, two challenges must be overcome: 1) a critical mass of adoption must be achieved, and 2) different systems

must be interoperable with one another. At a minimum, the data from these systems must be comparable. A paradox of some systems is that solutions that foster adoption may decrease interoperability, and vice-versa.

A key element in this conflict is the way terminology is used in health information systems. Standard terminology and structured data are a prerequisite for interoperability and sharing of medical information, decision-support and analysis. However, standard terminologies are unfamiliar and unnatural to most physicians, and as they struggle with applications that use them, they become frustrated. User adoption falls and errors can increase^[1]. It has been estimated that in the US, adoption of EHRs is only 9-17% in ambulatory settings and only 12-20% in inpatient settings.^{[ii],[iii]}

Can a highly usable terminology with links to reference terminologies and administrative classifications help solve both the problem of adoption and interoperability for both the developed and the developing world at once? We propose that it can and will introduce a program to test the hypothesis, called MGV-Net.^[iv]

1. Terminology about Terminology

Terminology is not just a list of names, or words, with numbers associated with them. A useful definition proposed by Chute which defines a terminology practically as: “that which enables users to invoke a set of controlled terms that correspond to formal concepts organized by a classification schema.”^[v] We can then further break down terminologies based on their characteristics and purposes. One useful differentiation is between administrative terminology, reference terminology, and interface terminology.

Administrative terminology is used primarily for the classification of information and the administration of health care delivery or reimbursement. Two examples of administrative terminologies are the International Classification of Diseases (ICD) and the Current Procedural Terminology (CPTTM). ICD is maintained by the World Health Organization and has become a standard for epidemiological reporting and administrative purposes. The U.S. and other governments maintain enhanced versions of ICD, called clinical modifications, containing greater detail to represent diagnoses. CPT is published by the American Medical Association and used for describing procedures. Neither is well-suited for capturing clinical data for EHRs, but both have been widely adopted in non-clinical settings.^[vi] A reference terminology is a concept-based, controlled medical terminology which allows for the complex organization and aggregation of clinical information. Examples include: the Systematized Nomenclature of Medicine–Clinical Terms (SNOMED[®]CT) and RxNorm.

Interface terminology, sometimes called *clinical interface terminology* or entry terminology, is used to describe lists of terms and phrases which are a “systematic collection of health care-related phrases (terms) that supports clinicians’ entry of patient-related information into computer program, such as clinical “note capture” and decision support tools.”^[vii] An example of interface terminology includes IMO’s Problem (IT)TM (previously known as Personal Health TerminologyTM).

2. What Is Different about Interface Terminologies?

In 1998, Cimino published his *Desiderata for Controlled Medical Vocabularies*^[viii] which was a seminal work outlining the requirements for a successful terminology. In 2005, in response to concerns that terminology needs to more closely mirror clinical

reality, he proposed additional desiderata. Whereas the original desiderata focused on the structure and content of controlled terminologies, the additional desiderata focus on the purpose and use of terminology (see Table 1).^[ix] This recognizes that how terminology is used is as important as what the terminology is. Reference terminologies do well for aggregation, reuse and inferencing, but suffer problems with properly capturing and retrieving information easily and accurately *with the least amount of information loss as possible*. These latter three desiderata require terminology more which clinicians use in everyday practice and which more closely approaches the level of specificity known at the point-of-care. This is the role of an interface terminology. An excellent review of interface terminology by Rosenbloom, et. al. provides an overview of the different requirements of interface and reference terminologies.^[vii]

Table 1: Additional Desiderata for Controlled Medical Terminology

Terminologies should capture what is known about the patient
Terminologies should support retrieval
Terminologies should allow storage, retrieval and transfer of information with as little information loss as possible
Terminologies should support aggregation of data
Terminologies should support reuse of data
Terminologies should support inferencing

Creating and managing a clinical terminology is difficult, particularly when you consider that terminology must support clinical applications, mapping to administrative and epidemiological coding schemes, and include multiple languages.^[x] Traditionally, the overarching design of terminologies included maps from reference terms to all other terms, including administrative classifications (see Figure 1 left). This was the most efficient method as different synonyms could then be mapped only once (to the reference term) and all subsequent maps would follow from the reference term. The difficulty arises when coding experts review the term-administrative code maps.

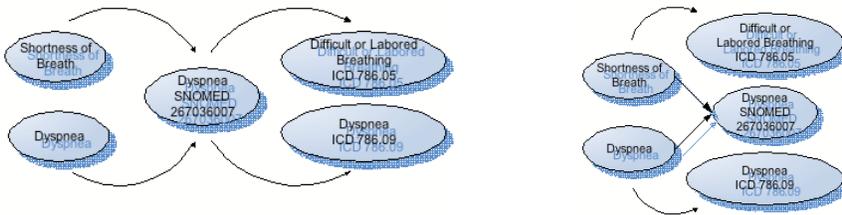


Figure 1: Traditional mapping between synonyms, reference terminology and administrative codes (left) Recommended organization with separate validated maps to administrative codes (right).

To ensure that coding is accurate, display terms should have explicit maps to administrative terms. For medico-legal reasons, the actual term chosen should have its mapping explicitly validated by a reputable organization like the American Health Information Management Association (AHIMA) (see Figure 1 right). This does not mean that interface terminologies should not be organized around concepts. It only suggests that the conceptual granularity of an interface terminology is often different than the reference terminology to which it is mapped. Interface terms still should have maps to concepts to ease updates and maintenance.

Interface terminologies should also be integrated with or have explicit maps to a reference terminology. Without the aggregation that is possible from a mapping to a reference terminology, there can be no consistent application of decision support and

analysis. The number of interface terms vastly complicates the visual representation of the data. It is much more feasible to write reports and decision-support rules against a parent reference term than all of the descendent reference terms plus.

However, reference terminologies frequently are missing concepts at relevant levels of clinical specificity. For example, SNOMED CT currently does not have reference terms for mild persistent asthma or mild intermittent asthma. However, these terms are in common use and are typically required for appropriate documentation and care of patients with asthma. Therefore, interface terms were created and mapped to the most relevant reference term(s). When more specific reference terms are subsequently added to SNOMED CT, the mappings from the interface terminology to SNOMED CT require updating, but terms used by clinicians (and front-end applications) would not have to change (see Figure 2).

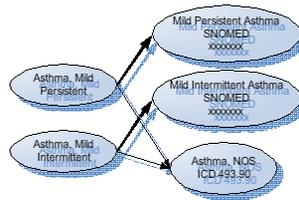


Figure 2: Remapping occurs when new reference terms are added.

3. Benefits of Interface Terminologies

Explicit interface terminologies such as PHT have been shown to markedly improve clinician satisfaction with EHRs in the US. The explicit mappings to reference terminologies have also advanced adoption of SNOMED CT. [xi] Centralizing the customization and maintenance of the interface terminology eliminates local terminology requirements and their resulting non-interoperable systems. The ability to rapidly add terms with the right level of clinical specificity satisfies user needs and allows application developers to implement systems and functionality quickly. In addition, common users of an interface terminology can share data using either interface or reference terms. Traditionally, site-specific terminologies exist as extensions to the reference terminology and remain site-specific pending addition of those terms to the reference terminology. They would remain site-specific permanently if the terms ultimately were not submitted or accepted.[xii] The usability benefits of interface terminologies when combined with the standards and structure of reference terminologies should significantly improve adoption of health information systems, particularly EHRs.

4. Health Information Systems in Africa: Unique Needs

It would seem obvious that the rationale for using appropriate terminology in health information systems applies to Africa as it does for the US and Europe. This would be true, but there are particular conditions in which terminology is even *more important* for implementation in Africa. Although Africa has large English-, French- and Portuguese-speaking groups, there are numerous linguistic differences at the country and local levels. Significant differences exist in the names of specific disease entities, medications, and laboratory tests. Although Ministry of Health reporting for individual

countries is frequently in one of the European languages mentioned above, local clinicians and providers who are recording the data may not speak these fluently. Conversely, clinicians trained in European languages may not be fluent in the languages used by their patients.

The gap between recording the information from the patient and what eventually arrives in the end as information, either locally or at ministerial level, is significant. Capturing knowledge in a manner in which it contains the most information is important, not only for the reasons propositioned by Cimino[ix], but also because much indigenous knowledge is being lost. Experience of health care providers in Africa is fragmented and knowledge is locked away in unanalyzed paper records and reports which take years to make their way to government where they are frequently ignored.[^{xiii}] Information technology should support the collecting and preservation of this knowledge, rather than suppress it.[^{xiv}] If terminology can help capture clinical experience accurately and quickly, it can then be put to use at the local and national level. Capturing the data using a method which incorporates reference terminology would make it possible for countries to share their knowledge of the status of health, disease and the effectiveness of interventions. In particular, the experience of neighboring countries would be more relevant to Africa providers and policy makers than data available from developed countries.

The need for information technology in the developing world is great. Effective implementation and application of health information technology will likely be required if any significant progress is to be made against the devastation of HIV/AIDS, tuberculosis and malaria.[^{xv}] The World Health Organization in its *World Health Report 2006: Working Together for Health*[^{xvi}] identified the shortage of trained health workers as one of the most important obstacles to strengthening health care systems in developing countries. As health care providers in Africa are frequently undertrained and overworked, health information systems can help extend their efforts by improving efficiency, and the quality of care can be improved through the use of appropriate guidance and information. There are many obstacles to successful design and implementation of these systems in Africa, which have been previously described. One challenge is that systems designed for US and European health facilities may not be applicable in rural and urban developing countries. Terminology has been identified as an important barrier[^{xvii}], and a properly implemented interface terminology has the potential to solve this problem. Taking evidence-based experience from other African situations and applying them within the information system holds promise for improving quality and helping to train providers in resource-poor settings.

However, the inclusion of medical terms in local languages, important for an interface terminology, is difficult and requires access to experts and reference materials.[^{xviii}] To test whether it is possible, the Millennium Villages™ Project (MVP) is building a network of health information systems called the Millennium Global Village Network (MGV-Net). MVP is the product of five years of intensive preparation by hundreds of scientists and development experts and works in eleven countries throughout Africa to help people lift themselves out of extreme poverty. MGV-Net uses a common data dictionary based on reference terminologies (SNOMED CT, LOINC and RxNorm), but employs a centralized Terminology Service Bureau (TSB) to manage an interface terminology distributed throughout the network.

The TSB is a critical component of the MGV-Net and maintains all terminology additions and mappings. To begin the process, a workbook identifying the most common diseases, diagnostic tests and medications was sent to all MVP villages for

translation. Clinicians in each of the MVP countries will work together online as part of the TSB to update the database.

5. Conclusions

We began this paper with a discussion about how interface terminologies are different from reference terminologies, and why interface terminologies are necessary for adoption of health information systems in the developed world. We ended by extending the value of these terminologies to Africa, and showed that it was even more important to the success of health information systems in developing countries. Capturing data in a culturally-sensitive manner, incorporating populations frequently missed by traditional information systems is important to an equitable health system and even a human right. ^[xix] The challenge is great for those working to alleviate poverty and improve health for those who most need it. We hope that health information systems which employ appropriate interface terminology to bridge the gaps between theory and reality will help overcome that challenge.

References

-
- ^[i] Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. *J Am Med Inform Assoc.* 2004 Mar-Apr;**11(2)**:104-12. Epub 2003 Nov 21.
 - ^[ii] Poon, et. al. Assessing the level of healthcare information technology adoption in the United States: a snapshot. *BMC Medical Informatics and Decision Making* 2006, 6:1 doi:10.1186/1472-6947-6-1
 - ^[iii] Burt CW and Sisk JE. Which Physicians And Practices Are Using Electronic Medical Records? *Health Affairs*, 2005 **24(5)**: 1334-1343 doi: 10.1377/hlthaff.24.5.1334
 - ^[iv] Kanter AS, et. al. Millennium Global Village-Net: Bringing together Millennium Villages throughout Africa. HELINA 2007, Bamako, Mali.
 - ^[v] Chute CG. Clinical classification and terminology: some history and current observations. *J Am Med Inform Assoc.* 2000 May-Jun;**7(3)**:298-303.
 - ^[vi] Cimino JJ. Review paper: coding systems in health care. *Methods Inf Med.* 1996 Dec;**35(4-5)**:273-84.
 - ^[vii] Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. *J Am Med Inform Assoc.* 2006 May-Jun;**13(3)**:277-88. Epub 2006 Feb 24.
 - ^[viii] Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods Inf Med.* 1998 Nov;**37(4-5)**:394-403.
 - ^[ix] Cimino JJ. In defense of the Desiderata. *J Biomed Inform.* 2006 Jun;**39(3)**:299-306. Epub 2005 Dec 9.
 - ^[x] Rector AL. Clinical Terminology: Why is it so hard? *Method Inform Med* 1999; **38**:239-52
 - ^[xi] Kanter AS, Cheriff A, Cole C, Cummins M, Alban C. Vocabulary Supporting Selection and Implementation to Enhance Physician EHR Experience. *Proc AMIA Symp*, 2005.
 - ^[xii] Dolin RH, et. al. Kaiser Permanente's Convergent Medical Terminology. *Medinfo.* 2004;**11(Pt 1)**:346-50.
 - ^[xiii] Gething PW, et. al. Improving imperfect data from health management information systems in Africa using space-time geostatistics. *PLoS Med.* 2006 Jun;**3(6)**:e271.
 - ^[xiv] Chikonzo A. The potential of information and communication technologies in collecting, preserving and disseminating indigenous knowledge in Africa. *Intl Info & Library Review.* 2006. **38**: 132-8.
 - ^[xv] Fraser, H, et. al. Implementing electronic medical record systems in developing countries. *Informatics in Primary Care* 2005;13-83-95.
 - ^[xvi] World Health Organization, *World Health Report 2006* (draft outline). http://www.who.int/hrh/whr06_consultation/en/index.html. Accessed February 6, 2008.
 - ^[xvii] Kamadjeu RM, Tapang EM, Moluh RN. Designing and implementing an electronic health record system in primary care practice in sub-Saharan Africa: a case study from Cameroon. *Inform Prim Care.* 2005;**13(3)**:179-86.
 - ^[xviii] Mbananga N, Mniki S, Oelofse A, Makapan S, Lubisi M. A model of developing medical terms in indigenous languages: a step towards consumer health informatics in South Africa. *Medinfo.* 2004;**11(Pt 2)**:1216-8.
 - ^[xix] Bambas L. Integrating equity into health information systems: a human rights approach to health and information. *PLoS Med.* 2005 Apr;**2(4)**:e102. Epub 2005 Apr 2

Cataloguing and displaying Web feeds from French language health sites: a Web 2.0 add-on to a health gateway

G. KERDELHUÉ^a, B. THIRION, B^a. DAHAMNA^a, SJ. DARMONI^a

^a *CISMeF, Rouen University Hospital, 76031 Rouen, France*

Abstract. Among the numerous new functionalities of the Internet, commonly called Web 2.0, Web syndication illustrates the trend for better and faster information sharing. Web feeds (a.k.a RSS feeds), which were used mostly on weblogs at first, are now also widely used in academic, scientific and institutional websites such as PubMed. As very few French language feeds were listed or catalogued in the Health field by the year of 2007, it was decided to implement them in the quality-controlled health gateway CISMeF ([French] acronym for Catalogue and Index of French Language Health Resources on the Internet). Furthermore, making full use of the nature of Web syndication, a Web feed aggregator was put online in to provide a dynamic news gateway called “*CISMeF actualités*” (<http://www.chu-rouen.fr/actualites/>). This article describes the process to retrieve and implement the Web feeds in the catalogue and how its terminology was adjusted to describe this new content. It also describes how the aggregator was put online and the features of this news gateway. *CISMeF actualités* was built accordingly to the editorial policy of CISMeF. Only a part of the Web feeds of the catalogue were included to display the most authoritative sources. Web feeds were also grouped by medical specialties and by countries using the prior indexing of websites with MeSH terms and the so-called metaterms. *CISMeF actualités* now displays 131 Web feeds across 40 different medical specialties, coming from 5 different countries. It is one example, among many, that static hypertext links can now easily and beneficially be completed, or replaced, by dynamic display of Web content using syndication feeds.

Keywords. Information Dissemination, Internet, Quality-controlled health gateway, Syndication feed

Introduction

The core of the World Wide Web is an extensive repository of documents linked by hypertext. It is now often referred to as “Web 1.0”. In this context, several quality-controlled health gateways have been developed [1]. These gateways were defined by Koch [2] as Internet services which apply a comprehensive set of quality measures to support systematic resource discovery. Their main goal is to provide a high quality of subject access through indexing resources using controlled vocabularies and by offering a deep classification structure for advanced searching and browsing. CISMeF ([French] acronym for Catalog and Index of French Language Health Resources on the

Internet) [1] was designed to catalog and index the most important and quality-controlled sources of institutional health information (N=36,103). Its URL is <http://www.chu-rouen.fr/cismef>.

Web 2.0 is a metaphor describing new Internet applications and services which emphasize greater user participation in developing and managing content [3]. As the use of Internet becomes more and more “social”, the need for bigger and faster information exchange grows. It manifests by the growth of Web syndication.

Web syndication is a form of information dissemination in which the (partial or complete) content of a website is made available for end-users or other sites to use. It is inherent to a large number of Web services, blogs, wikis, social networks, etc. The two main families of web syndication formats (XML-based) are RSS (Really Simple Syndication) and Atom [4]. These so-called Web feeds provide a way to rapidly disseminate awareness of new information. They permit continuous instant alerting to the latest ideas in medicine [5]. The adoption of the RSS format by PubMed to display search results is a remarkable example [6].

Convinced by the inherent qualities of Web feeds and of their interest for the medical domain, the CISMef team decided to catalogue health Web feeds available in French in January 2007. At that time, they were totally scattered across the Web. Some general lists and portals were existing but they mostly include Web feeds from journalistic or blogs sources. French language Web feeds from scientific, academic or institutional sources were neither listed nor catalogued in the health field.

The goal of this work is to describe how Web feeds were included in the existing catalogue and how a dynamic news gateway called *CISMef actualités* [7] was created.

1. Material and Methods

1.1. CISMef terminology

CISMef uses two standard tools to organize information: the MeSH thesaurus [8] from the US National Library of Medicine and several metadata element sets, in particular the Dublin Core metadata format [9]. In order to customize the MeSH to the field of health Internet resources, several enhancements [1] to this thesaurus were developed with the introduction of two new concepts, respectively metaterms (MT) and resource types (RT).

A metaterm denotes a medical specialty (e.g. *cardiology*), a biological science (e.g. *bacteriology*), or a health topic (e.g. *diagnosis*), which has semantic links with one or more MeSH terms, subheadings and RTs (N=123). The idea of creating metaterms came up to optimize information retrieval in CISMef and to cope with the relatively restrictive nature of these medical specialties as MeSH keywords. The metaterms are also used reciprocally to categorize a document from its specific indexing [10].

CISMef resource types (RT) are an extension of the publication types of MEDLINE. They are used to describe the nature of the resource, while MeSH terms describe its topic (N=275).

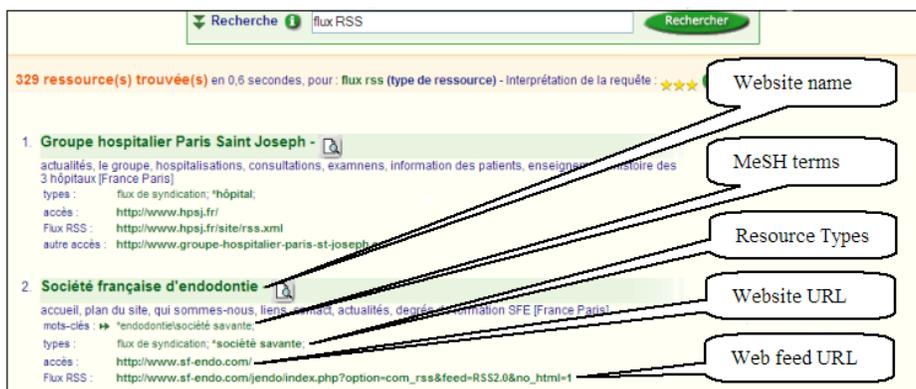


Figure 1: RSS Feeds as they appear in the catalogue after the query “RSS feeds”.

1.2. Implementing feeds in the catalogue

The first step was to retrieve the Web feeds. As nothing previously existed, it was done progressively and still it is an ongoing, not automated, process composed by several elements: survey of new or already indexed websites (progressively checked during day to day activities), use of search engines which permits to search Web feeds specifically.

The second step was to add two new elements: the web feed URL and a new RT (see Figure 1). As syndication feeds are always attached to a single website, it was decided not to create completely new entries in the catalogue but to add the feed description to the already existing entries for websites, thus attaching the prior indexing with MeSH terms to the feed itself. Two URLs are mentioned for each website providing web syndication: one for the website homepage, one for the web feed itself. The new RT created was called “syndication feed”, (“flux de syndication” in French). This resource type can be used in simple or advanced search to rapidly target the Web feeds among all the resources of the catalogue.

1.3. One step beyond the directory: displaying web feeds

Though the number of existing Web feeds is constantly growing, they remain unknown to a large part of the CISMéF audience. It was decided to display the feeds in order to make this content accessible for anyone without any prior technical knowledge.

Several free services offer to display a selection of Web feeds such as Google reader, Bloglines or Netvibes. However it was preferred to use an open-source software in order to have greater control over *CISMéF actualités* and to guarantee the independence and persistence of the service.

The Software chosen was Gegarius [11]. Its most deciding features were: to be completely web-based and running on a web server, to support multiple feeds formats, to provide a search engine, to be translated in French language, to be committed to Web standards (it renders XHTML/CSS), and finally to be a free software released under the *GNU General Public License*. The technical requirements were quite low: an Apache web server, PHP, and a MySQL Database. The software was then adapted to the CISMéF graphic charter by simply modifying its style sheet.

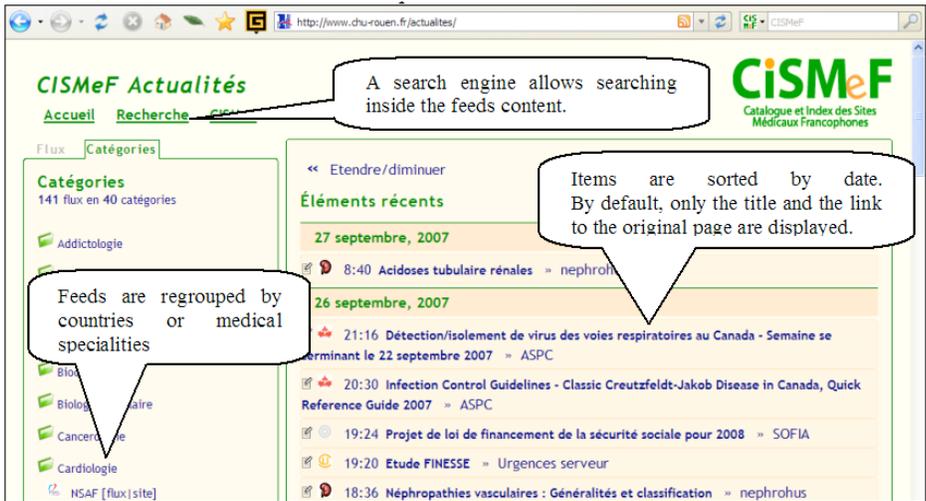


Figure 2. The CISMéF Actualités homepage.

1.4. Which feeds to display? Extending the CISMéF editorial policy to Web feeds

Not all the feeds catalogued in CISMéF are displayed in *CISMéF actualités* due to editorial choices but also to technical problems.

The CISMéF editorial policy is to reference only documents from the most authoritative sources within the large number of websites referenced. Thus, only the feeds provided by institutions, hospitals, medical schools and health professionals associations were displayed in *CISMéF Actualités*. On the other hand, feeds provided by the patients associations websites, the professional unions, though being indexed in the catalogue, were discarded.

We did not choose to display feeds from medical journals as there are still very few of them in French language and as their content is mostly not freely available.

The technical problems concern mostly Web feeds that do not comply with the RSS specification [12].

1.5. Categorize the Web feeds

Categorization is performed using semantic links between MeSH terms, MeSH qualifiers, CISMéF resource types and CISMéF metaterms (See Figure 2). For example, the web feed for the website CETL (URL: <http://www.cetl.net/>) indexed with the MeSH term “lysosomal storage diseases” was categorized in Genetics.

2. Results

2.1. Features of “CISMéF actualités”

The most important features of *CISMéF actualités* are:

- Almost real-time updates (while the CISMéF catalogue itself is updated weekly). Feeds are regularly checked for new items. As Gregarius itself does not have

scheduled tasks built in, a planned task using crontab executes every two hours the php script in charge of updating the feeds.

- Search engine. The content of each feed is accessible through full text search.
- Feeds are regrouped by countries and medical specialities (see Figure 2).
- Feeds aggregation and RSS output. For the whole CISMef actualités, for each category and for each search results, Web syndication is available.
- Name and Logo of the publisher clearly indicate the source of each item. For the CISMef team, the publisher is the main criterion of quality from the HON code of conduct [13].

2.2. Statistics

The number of Web feeds indexed in the catalogue is 329 (11 October 2007). 131 feeds are displayed in *CISMef Actualités* (11 October 2007). These feeds are categorized in 40 medical specialities. Thus 40 out of 123 (32%) CISMef metaterms are used. The number of unique visitors showed constant augmentation between July and September 2007 from 4675 to 7657.

2.3. Syndicate the aggregated feeds from CISMef Actualités

CISMef actualités provides a Web feed for each of these categories, aggregating the existing feeds. These new feeds can also be displayed in specialized context. We used this feature on a specialized gateway about handicap (<http://www.chu-rouen.fr/handicap>) where the feed of the category Handicap from *CISMef Actualités* is the core of the news section.

In the near future, we could implement a similar feature in a specialized search engine (<http://doccismef.chu-rouen.fr/servlets/KISMef>) in collaboration with the French National Cancer Institute (<http://www.e-cancer.fr/>).

3. Discussion

Integrating Web feeds in the CISMef gateway fulfils several needs. Users convinced by their usefulness can now find feeds in the French language medical domain through the catalogue itself. Users ignorant of Web feeds or lacking the technical knowledge to use them can now easily access their content through *CISMef Actualités*.

Very few papers in the medical literature refer to web syndication (n=15 in the MEDLINE bibliographic database September 27, 2007 with the query: "web syndication" or "RSS feed"). Most of them consist of explanations aimed at the end-users, very few concern the information providers [14] [15].

Two important websites provide service comparable to *CISMef Actualités*, though at a larger scale and concerning English language. They were major sources of inspiration.

The National Library for Health (UK) provides a large directory of RSS feeds accessible at this URL <http://www.library.nhs.uk/rss/Directory/> [16]. It shares with CISMef the human selection of the feeds, the ability to browse by categories and a search engine. But they do not display the feeds themselves and the search engine do not allow to search within the feed content.

Medworm (<http://www.medworm.com/>) [17] is a medical RSS feed provider as well as a search engine built on data collected from RSS feeds. It provides a search engine that search through the content of the feeds and sort the results by date and relevance. The content is also available through different categories.

The most important limit of *CISMeF Actualités* is the lack of Web feeds from major institutions. This is especially true for France. For example, The major Belgian and Canadian Health agencies provide Web feeds while the French don't.

The second limitation is the contents of the feed themselves. As they consist mostly of Press announcements, some major events in the Health field may not appear at all, while small events, like charity events for example, may appear out of range. *CISMeF Actualités* provides raw information, without the selection or the synthesis a journalist would do.

In the future, the growing number of websites deciding to provide Web feeds raises a number of questions. The discovery of Web feeds itself may need to be automated so that feeds from important sources would not be missed. Considering *CISMeF Actualités*, its content could be evaluated so that only the most important and relevant feeds are displayed.

References

- [1] Douyère M, Soualmia LF, Névéal A, Rogozan A, Dahamna B, Leroy JP, Thirion B, Darmoni SJ: Enhancing the MeSH thesaurus to retrieve French online health resources in a quality-controlled gateway. *Health Info Libr J* 2004 Dec: 21(4):253-61.
- [2] Koch T: Quality-controlled subject gateways: definitions, typologies, empirical overview, Subject gateways. *Online Information Review* 2000: 24(1): 24-34.
- [3] O'Reilly T. What is Web 2.0? Design patterns and business models for the next generation of software. <http://oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html> (accessed Jun 2007).
- [4] Web syndication - Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Web_syndication (accessed October 4, 2007).
- [5] Giustini D How Web 2.0 is changing medicine. *BMJ*, Vol. 333, No. 7582. (23 December 2006), pp. 1283-1284.
- [6] NLM Technical Bulletin, May-June 2005, RSS Feeds Available from PubMed. http://www.nlm.nih.gov/pubs/techbull/mj05/mj05_rss.html (accessed October 4, 2007).
- [7] CISMeF Actualités. <http://www.chu-rouen.fr/actualites/> (accessed October 4, 2007).
- [8] Medical Subject Headings - Home Page. <http://www.nlm.nih.gov/mesh/> (accessed October 4, 2007).
- [9] Dublin Core Metadata Initiative (DCMI). <http://dublincore.org/> (accessed October 4, 2007).
- [10] Darmoni SJ, Névéal A, Renard JM, Gehanno JF, Soualmia LF, Dahamna B, Thirion B. A MEDLINE categorization algorithm. *BMC Med Inform Decis Mak*. 2006 Feb 7;6(1):7
- [11] Gregarius - A Free, Web-based Feed Aggregator. <http://gregarius.net/> (accessed October 4, 2007).
- [12] RSS 2.0 Specification (RSS 2.0 at Harvard Law). <http://cyber.law.harvard.edu/rss/rss.html> (accessed October 4, 2007).
- [13] HONcode: Principles - Quality and trustworthy health information. <http://www.hon.ch/HONcode/Conduct.html> (accessed October 4, 2007).
- [14] Barsky E Introducing Web 2.0: RSS trends for health librarians *JCHLA / JABSC*, Vol. 27, No. 1. (Winter 2006), pp. 7-8.
- [15] J Robinson, S de Lusignan, P Kostkova, B Madge, A Marsh, C Biniaris The Primary Care Electronic Library: RSS feeds using SNOMED-CT indexing for dynamic content delivery. *Inform Prim Care*, Vol. 14, No. 4. (2006), pp. 247-252.
- [16] National Library for Health - RSS Directory. <http://www.library.nhs.uk/rss/Directory/> (accessed October 4, 2007).
- [17] MedWorm: Medicine RSS. <http://www.medworm.com/> (accessed October 4, 2007).

Videophones for the delivery of home healthcare in oncology

Mona LAILA^a, Vincent RIALLE^{a,b}, Lydie NICOLAS^c, Catherine DUGUAY^c and Alain FRANCO^{a,c}

^a*Laboratoire. TIMC-IMAG UMR CNRS 5525, Université Joseph Fourier, Grenoble France*

^b*Department d'Informatique Médicale, Centre Hospitalier Universitaire de Grenoble, Grenoble, France*

^c*Department de médecine aiguë et communautaire, Centre Hospitalier Universitaire de Grenoble, Grenoble, France*

Abstract. A videophone system was used to link cancer patients, undergoing chemotherapy at home, with care providers in the Home Healthcare Facility at the University Hospital Centre of Grenoble. The participant patients expressed their satisfaction both with the use and the technical quality of the system. Improvement was observed in the Hospitalisation Anxiety and Depression Scale (HADS), SF36 Health Survey Questionnaire and Palliative care Outpatient Scale (POS) scores during and at the end of the experiment. The results indicated that the use of videophones was both feasible and satisfactory, and that they may have a positive effect on the cancer patient's quality of life at home. Further studies are necessary to prove this final observation.

Keyword. Telemedicine, videophonic system, oncology, home health care, telehealth, acceptability, satisfaction, patient's quality of life

Introduction

Telemedicine provides new systems to deliver home healthcare for chronic diseases including cancer [1, 2].

Cancer patients often become anxious or even depressed a few months or years after the onset of their illness. The potentially devastating evolution of their disease along with the psychological suffering they experience gives rise to an increase in the number of hospitalizations, and consequently the consumption of health resources.

Several studies have evaluated the use of telemedical applications in oncology with the aim that health professionals share their sources of knowledge and exchange their experiences (expert opinions, imagery, pathological analysis of biopsies, radiotherapy, chemotherapy...) in order to improve cancer patient care [1, 3-12]. However, to our knowledge, no study has assessed the usability of videophones to deliver home healthcare in oncology.

Starting from this context, our study aimed to evaluate the feasibility and the utility of a videophonic system used for the delivery of health care to cancer patients undergoing chemotherapy at home.

1. Materials and methods

The study was carried out in the Home Healthcare Facility (Hospitalisation A Domicile - HAD), at Grenoble University Hospital (Centre Hospitalier Universitaire de Grenoble - GCHU), within the framework of the Medical Care Continuity project (MCC: <http://www.eten-mcc.org>).

1.1. Population

The study evaluated adult cancer patients receiving chemotherapy provided by the HAD facility at home.

The following patient inclusion criteria were used:

- Cancer sufferers receiving chemotherapy
- Age over 18
- A life expectancy of over 12 weeks.
- Show a vital stable state.
- Be receiving specialized medical care at home.
- Patient and general practitioner consent given prior to the experiment.

Patients who did not have adequate assistance at home and/or were unable to understand how to use the videophone system were excluded from the study.

1.2. Materials

A videophonic system was used to link each patient at home to the hospital staff. Videophonic communication was established by installing one videophone at the hospital (central work station – see figure 1) and a second one at the patient's home (patient work station – see figure 2).



Figure 1: central work station



Figure 2: patient work station

We evaluated the following items using standard scales and customized questionnaires as detailed below:

- Patient and patient family satisfaction and their acceptance of the use of the videophone. This was evaluated using three questionnaires each for the patient and for the family, all of which were prepared by the HAD facility team at GCHU.
- Patient anxiety and depression evaluated by the Hospital Anxiety and Depression Scale (HADS).
- Patient quality of life using the SF36 Health survey questionnaire.
- Quality of life and health care evaluated using the Palliative patient Outcome Scale (POS).

2. Results

Six patients, five women and one man, were included in this study. The median age of the patients was 66 years (minimum: 45, maximum: 76). Four patients were suffering from breast cancer, one from ovarian cancer and one from Myelodysplasia. The duration of participation in the study was three months for each patient. One patient died during the study and another was hospitalized for a long time following an exacerbation of his Myelodysplasia.

The HADS, the SF36 survey, the POS and specific questionnaires to evaluate patient and patient family satisfaction and acceptance of the videophone were completed 3 times during the study (at the beginning, middle and end). For the first patient studied, the POS and the HADS were not completed due to their later inclusion in the experiment.

2.1. The Patients' acceptance of the videophone

The participation in the study was proposed to twenty-two patients in the HAD facility at GCHU:

- Ten patients (45% of the total) agreed to participate in the experiment, among them the 6 patients recruited for the study. The four remaining patients could not be included for the following reasons:
 - Non-availability of a broadband internet connection.
 - No fixed phone at home.
 - Termination of their chemotherapy programme.
 - Termination of the study before being able to install the Videophonic equipment at the patient's home.
- Twelve patients (55% of the total) decided not to participate in this study owing to:
 - Feeling ill at ease with the technical machinery: two patients.
 - Lack of space at home: two patients.
 - Personal problems: two patients.
 - A feeling that such assistance was unnecessary given their good functional state: six patients.

2.2. The Patients' satisfaction

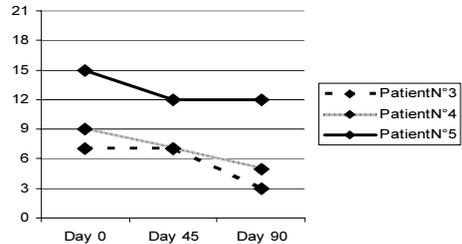
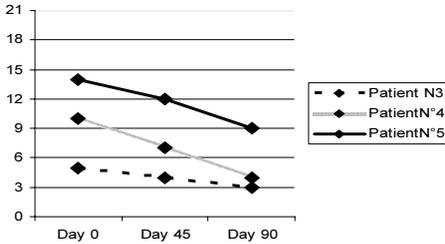
At the time of the initial presentation of the videophonic system, all the patients found its use simple and thought that it would accelerate and facilitate contact with carers and enable them to obtain a first medical opinion more rapidly.

No patient met with any difficulty in the use of the videophonic system during the study. Of those who continued, three found that the quality of the image was good, but not the quality of the sound. The other patient found that the quality of the image as well as the quality of the sound was good.

At the end of the experiment three of the four patients expressed their wish to keep the videophonic equipment

2.3. The Patients' anxiety and depression

The HADS scores were improved for the three patients who completed this scale (see graphs 1 and graph 2).

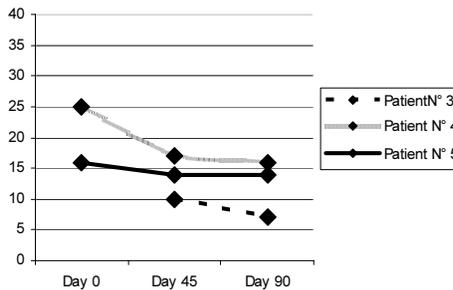


Graph 1. Evolution in Patients' Anxiety – beginning, middle and end of study

Graph 2. Evolution in patients' depression – beginning, middle and end of study

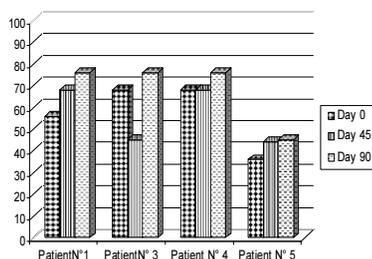
2.4. The Patients' quality of life

The three patients who completed the pos showed an improvement in its sum score by the end of the study (see graph 5).

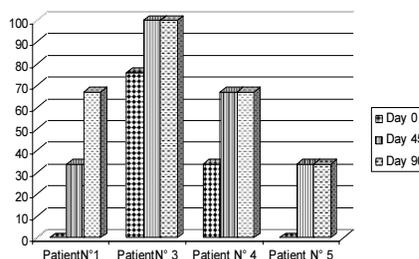


Graph 5. POS sum scores - patient version at the beginning, middle and end of the study.

The SF36 scores relating to emotional role and mental health improved for all the patients (see graph 3 and graph 4).



Graph 4: SF36 scores for mental health at the beginning, middle and end of the study.



Graph 3: SF36 scores for emotional role at the beginning, middle and end of the study.

3. Discussion

This study showed the feasibility and patients' acceptability of using videophones for the delivery of home healthcare in oncology [13-15].

The refusal to participate in this study was due to a rejection of videophones in two cases only. Six cases from the total refusals to participate were related to those patients' good general and functional state and their overall satisfaction with the home healthcare already delivered by the HAD facility without the use of videophones. The unease with technology, which was expressed by two patients in our study, is one of reasons for the refusal of technology not only by patients but also by caregivers [16-17].

The participating patients used the videophonic system to check their appointments at the hospital and to obtain explanations about the results of the different blood and radiological tests realised from their care providers. They also used it to check up on the doses of their new medicines. The apparatus was used by nurses visiting the patient's home to obtain the HAD facility doctor's agreement to administer chemotherapy in the case of the appearance of clinical symptoms, such as a rise in the patient's temperature or an alteration in blood results. The HAD facility doctors used the system to verify the position of the catheter used for giving chemotherapy in one patient, and to inspect the arm of another patient looking for a subcutaneous oedema.

The patients used the system successfully without any problems during all the duration of the experiment. At the end of the study they confirmed that the videophonic equipment had enhanced their feelings of security and their relations with the HAD facility doctors. They expressed their satisfaction relating to the technical quality of the system and the immediate answer obtained from the HAD facility doctors at each call.

The results showed a decrease in the patients' anxiety and depression levels exhibited by the falls in the HADS scores. The improvement in the SF36 scores for emotional role and mental health correlates with the decrease in the HADS scores. This result could be related to the positive effect of the videophonic system on the patients' anxiety and depression and on their state of mind, and in turn the effect of these on the

patients' quality of life and feelings of security[18]. This last is an observation that could not be proved in the study due to the small size of the sample.

In conclusion, while the use of videophones to deliver home healthcare seems to improve the cancer patient's quality of life, further studies are necessary to confirm their utility and effectiveness in the care of the cancer patient at home.

References

- [1] McAleer, J.J., D. O'Loan, and D.P. Hollywood, *Broadcast quality teleconferencing for oncology*. *Oncologist*, 2001. **6**(5): p. 459-62.
- [2] Tomuro, K., *Development of oral home telecare programme for the home-dwelling elderly: a pilot study*. *Gerodontology*, 2004. **21**(3): p. 177-80.
- [3] De Mello, A.N., et al., Development of a pilot telemedicine network for paediatric oncology in Brazil. *J Telemed Telecare*, 2005. **11 Suppl 2**: p. S16-8.
- [4] Doolittle, G.C. and A. Allen, Practising oncology via telemedicine. *J Telemed Telecare*, 1997. **3**(2): p. 63-70.
- [5] Norum, J. and M.S. Jordhoy, A university oncology department and a remote palliative care unit linked together by email and videoconferencing. *J Telemed Telecare*, 2006. **12**(2): p. 92-6.
- [6] Kristensen, I., et al., [Distance pediatric radiotherapy. Telemedicine a good tool to be used for discussions, exchange of experiences and competence]. *Lakartidningen*, 2006. **103**(15-16): p. 1188-90.
- [7] Karacapilidis, N. and D. Koukouras, A web-based system for supporting collaboration towards resolving oncology issues. *Oncol Rep*, 2006. **15 Spec no.**: p. 1101-7.
- [8] Gagliardi, A., et al., Feasibility study of multidisciplinary oncology rounds by videoconference for surgeons in remote locales. *BMC Med Inform Decis Mak*, 2003. **3**: p. 7.
- [9] Weinerman, B., et al., Can subspecialty cancer consultations be delivered to communities using modern technology?--A pilot study. *Telemed J E Health*, 2005. **11**(5): p. 608-15.
- [10] Stalfors, J., I. Bjorholt, and T. Westin, A cost analysis of participation via personal attendance versus telemedicine at a head and neck oncology multidisciplinary team meeting. *J Telemed Telecare*, 2005. **11**(4): p. 205-10.
- [11] Ogawa, Y., et al., Construction of a remote radiotherapy planning system. *Int J Clin Oncol*, 2005. **10**(1): p. 26-9.
- [12] Norum, J., et al., Telemedicine in radiotherapy: a study exploring remote treatment planning, supervision and economics. *J Telemed Telecare*, 2005. **11**(5): p. 245-50.
- [13] Demiriz, G., et al., Communication patterns and technical quality of virtual visits in home care. *J Telemed Telecare*, 2003. **9**(4): p. 210-5.
- [14] Passik, S.D., et al., A feasibility study of dignity psychotherapy delivered via telemedicine. *Palliat Support Care*, 2004. **2**(2): p. 149-55.
- [15] Clemensen, J., S.B. Larsen, and N. Ejksjaer, Telemedical treatment at home of diabetic foot ulcers. *J Telemed Telecare*, 2005. **11 Suppl 2**: p. S14-6.
- [16] Rialle, V., et al., What Do Family Caregivers of Alzheimer's disease Patients Desire in Smart Home Technologies? Contrasted results of a wide survey. *Methods of Information in Medicine*, 2008. **47**(1): P. 63-9.
- [17] Marshall, M., Technology and technophobia. *Journal of Dementia Care*, 2002. **10**(5): P. 14-5.
- [18] Montani, C., et al, "Telepsychometry": a remote psychometry consultation in clinical gerontology: preliminary study. *Telemed J*, 1996. **2**(2): P. 145-50.

Addressing the Biomedical Informatics Needs of a Microarray Laboratory in a Clinical Microbiology Context

Guillermo LOPEZ-CAMPOS, Victoria LOPEZ ALONSO¹, Fernando MARTIN-SANCHEZ

Medical Bioinformatics Department National Institute of Health “Carlos III”, Madrid, Spain

Abstract. For an effective integration of microarray-based technologies in clinical settings a number of contributions from biomedical informatics technologies and techniques are needed to facilitate the improvement of the phases of experimental design, image analysis, data management, annotation, and analysis. In this communication we briefly present the state-of-the-art in the application of biomedical informatics to laboratories conducting microarray experiments and how our unit is coping with these requirements imposed by the routine clinical work of the National Centre of Microbiology, a reference laboratory for the Spanish Health System.

Keywords: Biomedical informatics, microarray, bioinformatics clinical microbiology.

Introduction

Microarray technology is being integrated into basic biomedical research and is becoming a fundamental molecular monitoring tool in clinical microbiology settings, especially for diagnostic applications due to the possibility of multiplexing the simultaneous detection of several pathogens in a single reaction [1, 2, 3]. Most of these clinical applications are based in the development of DNA microarrays for the detection of specific genes or gene regions in pathogens, but others are based in protein (ELISA microarrays) or other immobilised molecules. Clinical microbiologists can also apply microarray technology from other approaches such as analysis of gene expression to monitor microbial metabolism and screening of regulons to study microbial response to drugs, environmental changes, genome organization, and evolutionary studies [4].

Many software development projects have emerged in response to the needs of postgenomic research projects focused on laboratory information management [5], scripting and programming language (bioperl, biopython, biojava, bioruby...) [6], database integration [7] and data models and ontology development [8].

Along with this increasing diversity of laboratory techniques and uses of microarrays there is an increasing need of specific biomedical informatics tools in some areas of the field of microarray applications in clinical microbiology that are not covered by the

¹Victoria López Alonso, Institute of Health “Carlos III”, victorialopez@isciii.es

solutions designed and developed for gene expression studies, mainly in cancer research.

The aim of this work is to identify some of the biomedical informatics needs detected in a microarray laboratory focused on clinical microbiology (Figure 1) and how these user requirements could benefit from different synergistic developments of bioinformatics and medical informatics techniques and methods [9].

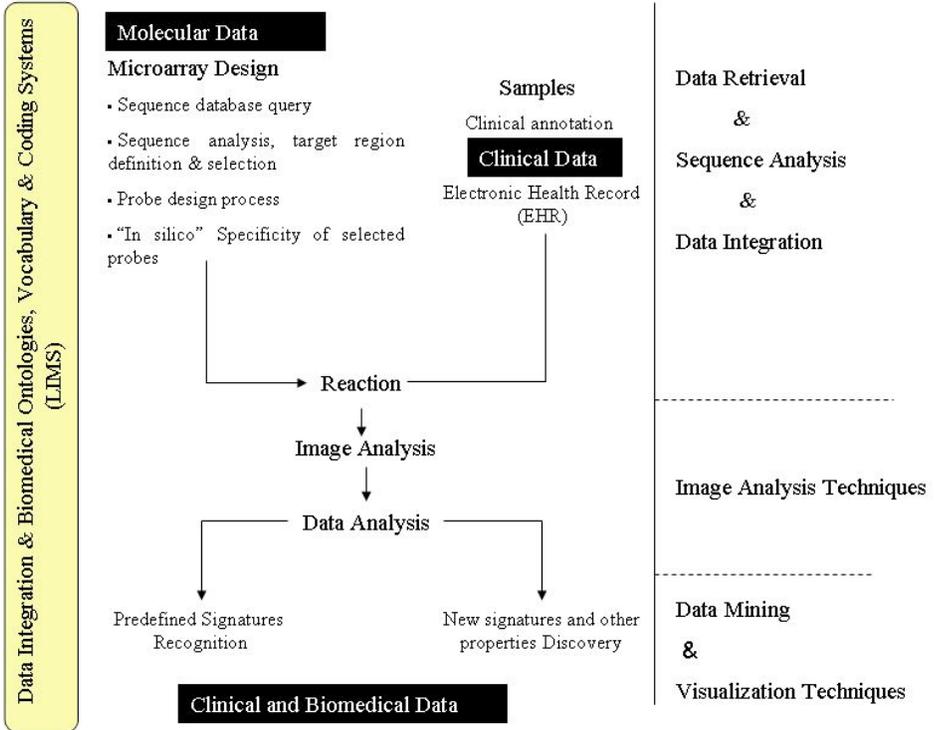


Figure 1. Biomedical informatics technologies of application in the different phases of a microarray experiment in a clinical microbiology context.

1. Material and Methods

1.1. Biomedical informatics and Microarray Design

Technological advances have boosted the field of microarray experiments. Although they represent high-throughput methods microarrays must be carefully designed in order to obtain useful data for clinical applications. Therefore design steps are key processes for the success or failure in the routine use of microarrays. In most of the clinical microbiological applications of microarrays, the design process is focused on the detection of probes that are capable to characterize and specifically identify the organism or organisms of interest. Therefore the initial step of design consists of

querying sequence databases and retrieving information. The target is the sequence of DNA to be amplified and fluorescently labelled. The strategy used to design it involves interrogating complete genome sequences stored in specific databases as the Comprehensive Microbial Genome [10] and expressed sequence tag databases available in the public domain. This step is problematic due in some cases to the vast amount and high redundancy of available sequence data or in other cases to the lack of available sequences in databases. At this point considerable effort has already been spent in generating EST libraries for the design of microbial genome microarrays.

Microarray probe design tools are very inefficient for large genomes and many existing tools operate in a batch mode. There are now several technologies that are useful for the design of microarrays for a selected set of model organisms based on Gene Ontology (GO) classification or using UniGene clusters [11].

Specificity of the designed oligonucleotides has to be checked with BLAST (Basic Local Alignment Search Tool) [12] and smart filtering techniques, which are employed to avoid redundant computation while maintaining accuracy.

1.2 Data management and annotation for Clinical Microarrays

Managing a microarray laboratory, especially for clinical use, typically involves good organization and sound tracking of experiments and samples data, including information on clinical samples, treatment, and experimental results. The use of Laboratory Information Management Systems (LIMS) addresses this problem by storing information with rigorous data entry mechanisms to prevent inaccuracies and to avoid redundancy. LIMS represent repositories of raw data associated with experiments, and are extremely useful as transient repositories before final submission of experimental microarray results to public databases, such as Gene Expression Omnibus (GEO) [13] or ArrayExpress [14].

1.3 Image Analysis Techniques in Microarray Experiments

In medical image processing the image content is often represented by features that are computed from the pixel matrix in order to extract features that support the development of reliable clinical diagnostic systems. In the case of microarray images there are some added difficulties for image analysis because many features are of abstract nature, as for instance those derived from a wavelength transform function.

After a reaction, microarrays are scanned and image analysis software determines the raw values using a data file that identifies the features and defines their dimensions and locations. Data stored in a tagged image file format (TIFF) is quantified by the image analysis software. To quantify microarray images it is necessary to ensure the quality of the signals and to prevent artefacts. Finally, the arithmetic mean or better median of the foreground and background pixels is calculated and data is obtained in the form of a table [15].

1.4 Microarray Data Mining Techniques

Before obtaining microarray results and applying sophisticated statistics, it is important to adequately normalize the data. A number of normalization methods have been proposed to correct systematic biases resulting basically from different amounts of

DNA used for labelling, different efficiencies of the Cy-3 and Cy-5 dyes in the labelling protocols, and different detection efficiencies of the dyes [16]. In gene expression experiments, fundamental patterns are extracted by several clustering methods like hierarchical clustering, self organizing maps and support vector machines. Lastly, the use of additional knowledge sources in the microarray data analysis process can improve the discovery and validation of clinical hypothesis. Information on sequence and structure, gene and protein interactions, function annotation and ontologies, or genetics and metabolic pathways can significantly complement any data analysis and improve its results.

2. Results

2.1. BUSSUB, a tool for designing clinical microarrays

The use of oligonucleotide microarray technology requires a very detailed attention to the design of specific probes to be spotted on the solid phase. They must have high sensitivity to detect a target gene and must be unique with high specificity. Parameters such as probe length, number of oligonucleotides, maximal distance from the 3 end, melting temperature range, threshold to reject secondary structures and prohibited sequences have to be considered in order to obtain an efficient design. In addition, the designed oligonucleotides must be computationally optimized to achieve greater specificity and uniformity and to perform optimally under the same melting temperature and other experimental conditions, reducing the noise due to cross-hybridization. Although there is an increasing number of publicly available genetic and molecular analysis programs for microarray oligonucleotide design, most of them use the same algorithm or criteria with only scarce variation [17]. None of them gathers the whole set of criteria. For this reason we have developed BUSSUB [3] an amplicon retrieval software that simplify and boost the process of recovering sequences contained between two given regions. BUSSUB is essential is the first stage of microarray design because this system is able to deal with high complexity files including several complete bacterial genomes.

2.2. AMANDA, Clinical Microarray Laboratory Information Management System

We have developed AMANDA [2] a LIMS specifically designed to store information about patient data, sample description, experimental resources, experimental parameters and conditions, and raw and processed hybridization results. The integration of this information into the microarray information system facilitates the processes of quality control and assessment of microarray experiments as well as further clinical interpretation of the results. AMANDA is compliant with the minimal information about a microarray experiment (MIAME) format [18]. This system is able to store, manage and visualize all the data and the information coming from several studies, to process it as a homogeneous dataset instead of multiple and separate sources

2.3 Image Analysis Software

Image analysis software used in our microarray laboratory carries out three fundamental tasks: gridding (to locate each spot on the slide), segmentation (to

differentiate the pixels within a spot-containing region into foreground) and information extraction. Information extraction includes two steps: spot intensity extraction and background intensity extraction. Spot intensity extraction refers to the calculation of fluorescent signal of the foreground from segmentation process, while background intensity extraction utilizes different algorithms to estimate the background signal due to the non-specific hybridization on the glass.

Developments in image acquisition, analysis and informatics technologies are ongoing and are expected to broaden the usefulness of DNA microarrays. For example Stanford MicroArray Database (SMD) [19] has the ability to store, retrieve, display and analyze the complete raw data produced by several additional microarray platforms and image analysis software.

2.4. Information of Clinical Microarray experiments

Most of the information resulting from diagnostic microarray experiments in a clinical microbiology context is qualitative (presence or absence of signal). In other cases as in gene expression or protein assays, data need to be analyzed in the context of clinical and epidemiological information in order to extract relevant knowledge useful for developing clinical solutions (Figure 2).

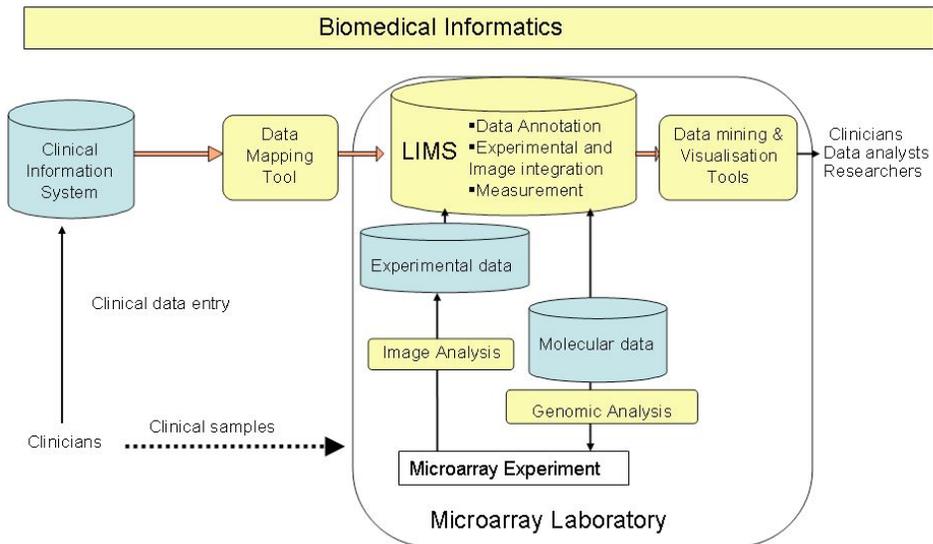


Figure 2. Implementation of biomedical informatics tools required for the production and use of clinical microarrays.

Data arising from microarray experiments can be normalized, filtered and analyzed utilizing several normalization and statistical modules available on integrative web based services. SMD [19] and the Bioconductor project [20] have added several tools and new ontologies to allow for an accurate and searchable annotation of biological

samples and experiments, developing and generalizing the schema for more efficient and flexible storage and analysis of microarray data.

Conclusions

High-throughput technologies such as microarrays are in the process of revolutionizing clinical microbiology. Further research and development of new biomedical informatics methods and techniques is becoming an integral part of a microarray laboratory. In our experience, these technologies are key in the process of assigning experimental genomic signatures to clinical profiles and to turn data into meaningful and reproducible clinical and mechanistic inferences.

References

- [1] A. Loury and L. Bodrossy, "Highly parallel microbial diagnostics using oligonucleotide microarrays", *Clinica Chimica Acta*, vol. 363, pp. 106-119, 2006.
- [2] G. López-Campos, L. Garcia-Albert, F. Martín Sanchez and A., Garcia-Saiz, "Analysis and management of HIV peptide microarray experiments", *Methods of Information in Medicine*, vol 45, pp. 158-162, 2006.
- [3] G. López-Campos, M. Coiras, J.P. Sánchez-Merino, M.R. López-Huertas, I. Spiteri, F. Martín-Sanchez, P. Pérez-Breña, "Oligonucleotide microarray design for detection and serotyping of human respiratory adenoviruses by using a virtual amplicon retrieval software" *Journal of Virological Methods*, vol. 145, pp. 127-136, 2007.
- [4] A. Ehrenreich, "DNA microarray technology for the microbiologist: an overview" *Applied Microbiology and Biotechnology* vol. 73, pp. 255-273, 2006.
- [5] G. Thallinger, S. Trajaneski, G. Stocker and Z. Trajanoski, "Information management systems for pharmacogenomics", *Pharmacogenomics*, vol 3 pp. 651-667, 2002.
- [6] <http://www.open-bio.org>
- [7] V. Maojo, M. Garcia-Remesal, H. Billhardt, R. Alonso-Calvo, D. Perez-Rey and F. Martín-Sanchez, "Designing new methodologies for integrating biomedical information in clinical trials", *Methods of Information in Medicine*, vol 45, pp 180-185, 2006.
- [8] <http://www.obofoundry.org/ontologies.shtml>
- [9] D. Rebholz-Schuhman, G. Cameron, D. Clark, E. van Mulligen, J.L. Coatrieux, E. Del Hoyo Barbolla, F. Martín-Sanchez, L. Milanesi, I. Porro, F. Beltrame, I. Tollis and J. Van del Lei, "SYMBIOmatics: Synergies in Medical Informatics and Bioinformatics-exploring current scientific literature for emerging topics", *BMC Bioinformatics*, vol 8, (Suppl 1) S18, 2007.
- [10] <http://cmr.tigr.org/tigr-scripts/CMR/CmrHomePage.cgi>
- [11] "Onto-Tools, the toolkit of the modern biologist: Onto-Express, Onto-Compare, Onto-Design and Onto-Translate", *Nucleic Acids Research*. vol 31, pp 3775-3781, 2003.
- [12] <http://www.ncbi.nlm.nih.gov/BLAST/>
- [13] T. Barret and R. Edgar, "Mining microarray data at NCBI's Gene Expression Omnibus (GEO)", *Methods of Molecular Biology*, vol. 338, pp. 175-190, 2006.
- [14] H. Parkinson, M. Kapushesky, M. Shojatalab, N. Abeygunawardena, R. Coulson, A. Farne, E. Holloway, N. Kolesnykov, P. Lilja, M. Lukk, R. Mani, T. Rayner, A. Sharma, E. William, U. Sarkans and A. Brazma, "ArrayExpress-a public database of microarray experiments and gene expression profiles", *Nucleic Acids Research*, vol. 35, pp. 747-750.
- [15] A. Petrov, S. Shah, S. Draghici, S. Shams, " Microarray image processing and quality control" in S. Shah, G. Kamberova (eds) DNA array image analysis- nuts & bolts. DNA, Eagleville, pp 99-130
- [16] J. Quackenbush, "Microarray data normalization ", *Nature Genetics*, vol 32, pp 496-501
- [17] W. Rychlik "OLIGO7 " *Methods of Molecular Biology*, vol. 402, pp 35-60, 2007.
- [18] A. Brazma, P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoecker, J. Aach, W. Ansorge, C.A. Ball, H.C. Causton et al. "Minimum information about a microarray experiment (MIAME) - toward standards for microarray data", *Nature Genetics*, vol 29, pp 365-371, 2001.
- [19] <http://genome-www5.stanford.edu/resources/restech.shtml>
- [20] <http://www.bioconductor.org>

CEMARA: a Web Dynamic Application Within a N-tier Architecture for Rare Diseases

Claude MESSIAEN¹, Loïc LE MIGNOT¹, Ana RATH³, Jean-Baptiste RICHARD¹,
Eric DUFOUR², Mohamed BEN SAID¹, Jean-Philippe JAIS¹, Alain VERLOES⁴,
Martine LE MERRER⁵, Christine BODEMER⁶, Geneviève BAUJAT⁵, Marion
GERARD-BLANLUET⁴, Eva BOURDON-LANOY⁶, Rémi SALOMON⁷, Ségolène
AYME³, and Paul LANDAIS^{1*} for the CEMARA task force

¹ *Paris Descartes University, UPPRESS EA 4067 and Service de Biostatistique et
d'Informatique Médicale, APHP - Necker-Enfants Malades Hospital*; ²*SELIC;*
³*Orphanet and INSERM SC 11*; ⁴*Department of Genetics, APHP- Robert DEBRE
Hospital*; ⁵*Department of Genetics*, ⁶*Dermatology unit and* ⁷*Paediatric Nephrology
unit, Necker-Enfants Malades Hospital, Paris - France*

Abstract. Rare diseases include a group of conditions characterized by a prevalence lower than 5 per 10,000 in the community. In France, any rare disease affects less than 30,000 patients and often much less. Three to 4% of children and 6% of the population in Europe are affected. It is a true public health stake since most diseases do not have any curative treatment. In France, the Ministry of Health has initiated a National Rare Diseases Plan. Twenty five out of 132 labelled Reference Centres (RC) decided to share a common Information System named CEMARA. It is dedicated to collect continuous and complete records of all patients presenting with a rare disease, and their follow-up. The main objective of CEMARA is to contribute to the missions of the RC regarding the registration and description of their activities, coordination of the network of their correspondents, organization of the follow-up of rare diseases, and analysis of the epidemiological patterns. A description of CEMARA is provided as well as its cooperation with Orphanet and Genatlas, and a presentation of 11803 current records collected by more than 300 health care professionals belonging to more than 70 sites.

Keywords. Dynamic Web interface; Decisional Information System; Thesaurus; Rare Diseases;

Introduction

Rare diseases encompass a group of conditions characterized by a low prevalence. In Europe, a disease is "rare" when less than one out of 2,000 people suffers from a

* Corresponding author : Pr Paul LANDAIS – Service de Biostatistique et d'Informatique Médicale – Hôpital Necker-Enfants Malades 149 rue de Sèvres 75743 Paris, France email : landais@necker.fr

specific clinical syndrome. In France, rare diseases might affect 30,000 patients. For the European Community, it means up to 230,000 patients. According to the definition, 5000 to 7000 rare diseases are currently identified. Three to 4% of children and 6% of the population in Europe is affected. It is a true public health concern since most diseases do not have any curative treatment. In France, the Ministry of Health has initiated a National Plan for Rare Diseases; 132 Reference Centres (RC) [1] have been labelled for a group of diseases or, even a given disease such as “constitutional bone disorders” or “juvenile arthritis”, respectively. Among their missions, the RC are involved in the epidemiological monitoring of pathologies they are in charge of, in order to better assess their distribution on the territory, the existing management networks, as well as social, educational and familial repercussions of these pathologies. An Information System named CEMARA (CEntres MALadies RAres) was set up in order to collect continuous and complete records of all patients presenting with a rare disease, and when applicable, their follow-up.

The main objective of CEMARA is to contribute to the missions of the RC regarding the registration and description of their activities, coordination of the network of their correspondents, organization of the follow-up of rare diseases, and analysis of the epidemiological patterns.

1. Material and Methods

1.1. CEMARA professional network

Presently, 25 centres are members of the CEMARA network in France. Each centre is composed of one or more clinical units, managed by a coordinator and distributed on the whole French territory. A medical correspondent is designated for each site (a map is available at <https://cemara.org/presentation/show.jsp?sm=membres>). Close to 300 members are registered for 70 sites. All gave their consent and signed the confidentiality charter of the network. CEMARA obtained its agreement in accordance with the French Protection Act.

1.2. Thesauri: conception and constitution

For the diagnosis labels of the patient records, a classification was set up with the health professionals corresponding to their needs. For rare diseases, the International Classification of Diseases 10th version (ICD10) does not provide the necessary support since its granularity is not appropriately designed for such a target, and many rare genetic disorders are not listed in. Moreover, the structure of ICD10 is not adapted to a field where nosology evolves very rapidly (for example following new genetic discoveries leading to splitting existing clinical diagnoses) and where 300 new syndromes are reported every year.

We thus linked our work to Orphanet [2]¹ that designed specific thesauri dedicated to rare diseases. Orphanet has a know-how in the domain and is member of the European Rare Disease Task Force [3]. It provides an ontological support to experts for each field of CEMARA. The specificity of this collaboration provides a shared ontology for each thesaurus of CEMARA with a lexical uniqueness for the labelling of each disease including the management of synonyms and eponyms associated to some diseases or disorders (for instance De Toni-Debré-Fanconi syndrome and proximal

tubulopathy). Furthermore, the associated nosography/nosology helps health professionals from different fields to share common terms. Orphanet can also benefit from a feedback through CEMARA by studying, at steady rate, the associations of diagnosis with keywords and/or notes entered by the physicians, which can lead to evolutions of the thesaurus. Regarding gene mutations we use GENATLAS, which contains relevant information with respect to gene mapping and genetic diseases [4]. A specific link was also developed towards this database. OMIM codes will be also provided.

1.3. Information System (IS): Design and Implementation

The IS of CEMARA collates in a standardized representation a minimal patient record elaborated by paediatric and adult health professionals. CEMARA aimed at fulfilling several requirements: scalability, portability, reliability, accessibility and cost effectiveness oriented toward non-proprietary software. The architecture of CEMARA is based on a n-tier architecture. Via a web browser, the client tier connects to the middle tier, which is connected to several databases: the production database, the geographical dictionary database, and the thesaurus database. A data warehouse and a geographical information system allowing queries and representation are in progress. Their framework is close to an already available application for end-stage renal disease [5]. The middle tier supports client services through Web containers and business logic services through component containers. Business logic in the middleware is interfaced with a SGBD dependant handler which supports the transactions toward the database. At the client side, CEMARA relies on existing local Internet networking facilities and on a widely spread computer configuration in medical settings. Access via a personal digital assistant is also available.

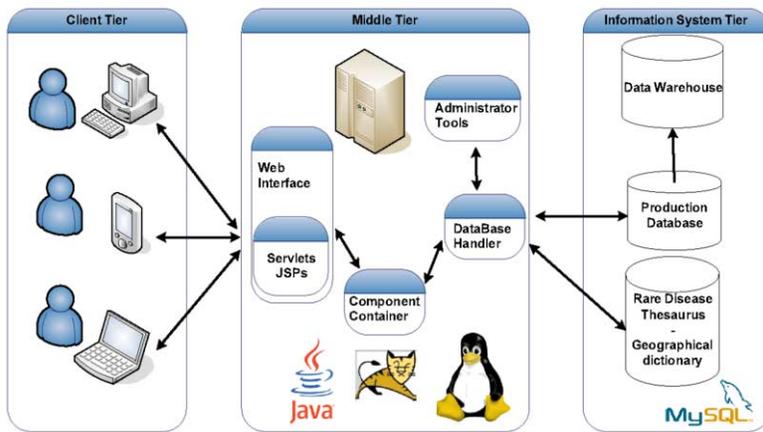


Figure 1. N-tier architecture for CEMARA

Maintenance and evolutions are made centrally, which reduce deployment costs and delays. The structure of the production database matches the structure of the CEMARA network and respects privacy, confidentiality and security [6]. A specifically developed double-entry function prevents identity doubles [7].

The setting-up began on May 23rd, 2007, initiated by fifteen centres. Coordinators and users were trained, through centralized and on-site sessions. A clinical research assistant organized the training and technical assistance.

CEMARA also includes institutional collaborations with: the Ministry of Health, the National Institute for Health Surveillance, the “Haute Autorité de Santé” and the National Health Insurance Fund concerning the design of national protocols for diagnoses, treatment and follow-up, as well as for cost analyses, associations of patient families for improving patient care, university hospitals, for the coordination and connection of professionals through the CEMARA network.

1.4. Interoperability

CEMARA was conceived in order to communicate with other sources of information: the use of XML, as an exchange format, permits a greater flexibility and better capacities to exchange data with other information systems such as French Medical Insurance system or Hospitals Information systems. It also allows importation of former databases (Figure 2).

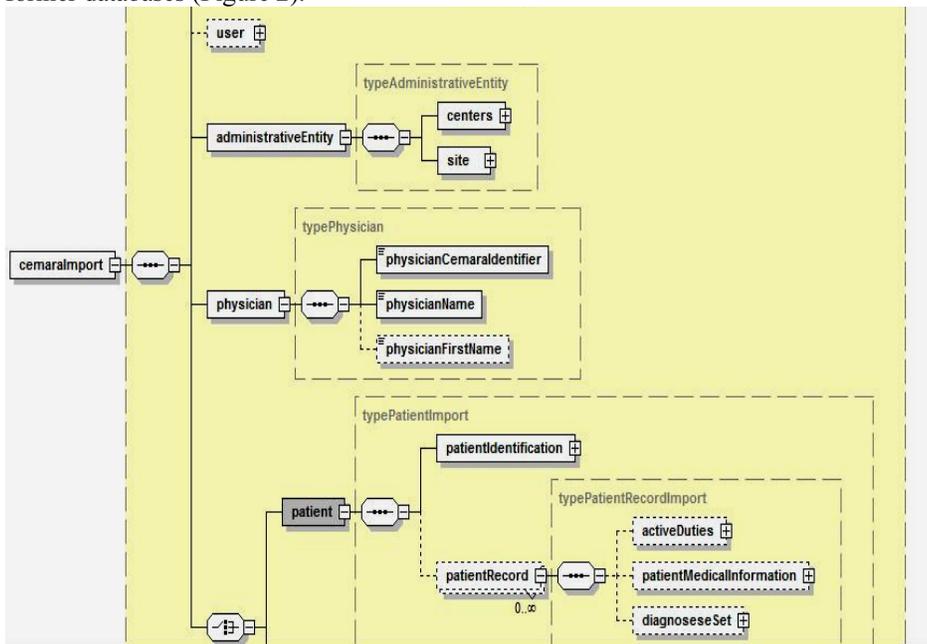


Figure 2. CEMARA: Part of an XML data import scheme

1.5. Data entry sheets

All the centres share a common core of information that includes identification data (for the index case and relevant family members), diagnosis, context and medical activity. Patient identification is based on: name, surname, birth date, death date, place of residence... Modalities of medical activities and modes of recruitment are recorded. Optional information is gathered about the patient's medical history e.g. antenatal and/or neonatal information. One or more diagnoses are labelled using the corresponding thesaurus, as well as the status of this diagnosis (probable, confirmed, on, in process) a chromosomal description of the anomaly if required, and additional

keywords for atypical signs and symptoms, or for the patients with an unknown diagnosis. Keywords were prepared by the steering committee. Diagnosis selection use Asynchronous JavaScript And XML (AJAX) queries. Complementary information for specific diseases are described in the so called “petals”, attached to the core data, and focused on specific data collection.

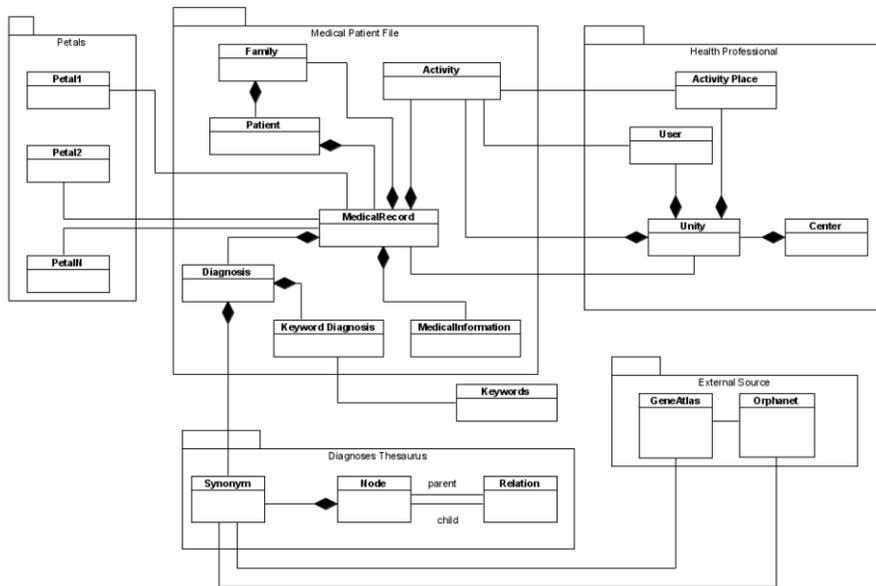


Figure 3 – UML scheme of CEMARA

2. Results

More than 300 health care professionals belonging to more than 70 sites of the 25 RC contributed to the data collection. Currently, the data set gathers 11,803 records and includes 6076 male patients (340 foetuses) and 5536 females (345 foetuses). For 191 foetuses gender was non determinable. Median age was 6.8 years. Patients with rare diseases were mainly referred to centres by paediatricians (4168) and medical specialists (2918), while for foetus by gynaecologists (612) and by centres for prenatal diagnosis (273). These patients were mainly enrolled via outpatient clinics (9134) or hospital wards (689). Foetal cases were described either during pregnancy (519), or after spontaneous miscarriage or termination of pregnancy (446). As expected, the number of patients per diagnosis is limited, even for the most prevalent diagnoses. Moreover, roughly 1/3 records are classified as “unknown diagnosis”, meaning that no diagnosis has been reached, which is a common occurrence for patients with multiple malformations, whose syndrome do not fit with currently delineated entities.

3. Discussion

The EC program for rare disease, initiated in 2003, aimed at improving information and knowledge for the development of public health [8]. In France, the rare diseases plan played a key role in promoting the management of these diseases [9]. The RCs have to standardize their information and structure its collection. It triggered the development of CEMARA. The experience shared with Orphanet and Genatlas enabled us to extend the expertise gathered with the French national REIN program [2] we developed for end-stage renal disease. It strengthens the durability of our application. Furthermore, the application is based on open source software, which allows lower development costs. The conception of an evolutionary design conveyed by its modularity ensures its scalability. Within nine months, close to 12,000 rare disease cases were recorded. It shows the involvement of the professionals and the utility of this application. Users of databases often claim that they don't have a real feedback on the data they feed. Therefore, we provide every three months reports on activity and diagnoses. Each participant can use a multicriteria search tool to explore his case-mix. We also prepared advance epidemiological analysis computed with R [10] on anonymous and aggregated data using our experience derived from the SIGNE, a web geographical information system [11]. CEMARA allows assessing the geographical distribution of existing cases for providing the appropriate offer of care. It allows better describing new cases of rare diseases and organizing their follow-up, and identifying the medical and psycho-social needs. The use of a web-based interface intends to help medical recording, management and reporting of rare diseases. It aims to contributing to the public health adaptation to the offer of care with a better knowledge of diseases, thus providing a support to health care decision-making in this complex and specific domain.

Acknowledgments

The coordinators of the 25 CEMARA centres are warmly acknowledged for their participation to the program as well as the members of all the centres

References

- [1] http://www.alliance-maladies-rares.org/IMG/pdf/Liste_Centres_de_Reference.pdf
- [2] CEMARA et Orphanet : une collaboration au service des centres de référence, <http://www.orpha.net/actor/Orphanews/2007/070321.html> [last view Feb 28, 2008]
- [3] Rare Disease Task Force : <http://www.rdtf.org/> [last view Feb 28, 2008]
- [4] GENATLAS <http://www.dsi.univ-paris5.fr/genatlas/> [last view Feb 28, 2008]
- [5] Landais P, Simonet A, Guillon D, Jacquelinet C, Ben Saïd M, Mugnier C, Simonet M. SIMS REIN: a multi-source information system for end-stage renal disease. *C R Biol.* **325** (2002), 515-28.
- [6] Rindfleish TC. Privacy, Information Technology and Health Care. *Communications of the ACM* August **40** (1997), 93-100
- [7] Le Mignot L, Mugnier C, Ben Saïd M, Jais JP, Richard JB, Le Bihan-Benjamin C, Taupin P, Landais P. Avoiding doubles in distributed nominative medical databases: optimization of the Needleman and Wunsch algorithm. *Stud Health Technol Inform.* **116** (2005), 83-8.
- [8] http://ec.europa.eu/health/ph_projects/2004/action1/action1_2004_10_en.htm [last view Feb 28,2008]
- [9] http://www.sante.gouv.fr/hm/actu/maladie_rare/plan.pdf
- [10] The R Project for Statistical Computing: <http://www.r-project.org/> [last view Feb 28, 2008]
- [11] Richard JB, Toubiana L, Le Mignot L, Ben Saïd M, Mugnier C, Le Bihan-Benjamin C, Jais JP, Landais P. A Web-based GIS for health care decision-support. *AMIA Annu Symp Proc.* (2005), 365-9.

Developing a Shared Electronic Health Record for Patients and Clinicians

Cornelia M. RULAND,^{ab} Haakon BRYNHI,^c Roar ANDERSEN,^a Tore BRYHNI^d

*^aCenter for Shared Decision Making and Nursing Research,
Rikshospitalet Medical Center, Oslo, Norway,*

^bDepartment of Medicine, University of Oslo,

^cDepartment of Computer Science, University of Oslo ^dFaster Imaging AS

Abstract. Improving Shared Decision Making (SDM) and patient-provider communication, and providing all citizens with equal access to health information has become a high priority health policy goal. In this interdisciplinary, international research collaboration we develop and test *CONNECT* (Care Online: Novel Networks to Enhance Communication and Treatment), a patient portal that integrates a suite of context-sensitive patient communication and information tools into a patient-clinician shared electronic health record that patients can use seamlessly through heterogeneous networks from different locations (home, hospital, doctor's office). In this paper we present methods used to develop *CONNECT*; how to safeguard data security and confidentiality and adapt user interfaces to different users, devices and contexts of use; as well as ensure safe and efficient data transfer through heterogeneous networks; and critical success factors and challenges.

Keywords: Electronic Health Record, Patient Portal, Mobile Solutions, Internet Support

Introduction

A rapidly growing amount of literature has addressed the importance of Information and Communication Technology (ICT) to improve SDM and communication between health care providers and patients. In many countries, empowering patients to take a more active role in their own health and provide all patients with equal access to health services and information have been declared high priority health policy goals[1;2] Web-based and other innovative ICTs have shown great potential to provide patients with information and communication specific to their needs. To utilize their full potential, such ICTs need to be integrated into the practice of regular health care. Personal and Shared Health Records have therefore, been declared an international research priority because of their potential to provide an environment in which health information about an individual can flow seamlessly among systems used by authorized health professionals, care givers and the patient[3]. In Norway, a number of public policy documents[1;2;4] outline strategies for extended use of the EHR.

To meet these goals, the Center for Shared Decision making and Nursing Research at Rikshospitalet Medical Center, with funding from the Norwegian Research Council, is currently developing *CONNECT* (Care Online: Novel Networks to Enhance Communication and Treatment). *CONNECT* is a patient portal that integrates a suite of context-sensitive, patient communication and information tools into a patient-clinician shared EHR to support shared decision making, patient-provider communication and patient-centered care. Developed for different environments (hospital, outpatient clinic and patients homes), these tools include (1) the *Choice* application used at the point of care, where patients on a tablet PC report and share data about their symptoms and illness experiences with their care providers to support shared care planning; (2) *WebChoice* an Internet-based support system that cancer patients can access from home through a secure patient portal. (3) In addition, we develop *CONNECT* mobile, a mobile client with device adaptation, providing mobile access to *WebChoice* information with seamless and secure network access.

The natural next step is to combine these tools into an integrated application within the context of a shared EHR. This can provide a seamless environment where patients can share their data, communicate with their care providers, and obtain support from where ever and whenever they need it, all within the same system. While the suite of *Choice* tools so far have been stand-alone systems primarily used by cancer patients, the underlying principles are generic and transferable to other patient groups. Therefore, we integrate these applications into regular health care which holds great promise to improve patient-provider relationships, patient self-management, effectiveness and continuity of care, and consequently patient safety for a wide range of health care users.

In this paper we present (1) methods for integrating the *Choice/WebChoice* applications into the EHR; and (2) *CONNECT* mobile, that provides a mobile solution of *CONNECT* in a manner that safeguard data security and confidentiality and adjust user interfaces to different users, devices and contexts of use.

1. Integrating Choice and WebChoice with the EHR

Currently *WebChoice* is a self-contained Web application that patients use from home through a standard Internet browser. It contains:

1. *Choice*, the comprehensive *Assessment tool* that allows patients to report and monitor symptoms, problems and priorities for care along physical, functional and psychosocial dimensions, currently and over time. Using branching logic, questions are personalized to previous responses so that only relevant questions are asked. It also includes ratings for “degree of bother” and “priorities for care” to assist provider and patient in rapidly determining which areas are most important to address. The system creates immediately a summary report that displays patients’ reported symptoms ranked by their priorities for care.
2. *Tailored Symptom Self-management Support*. At any time, the patient can obtain helpful advice from the Knowledge Base about how to self-manage their symptoms. Patients’ self-reported symptoms trigger the display of the appropriate subset of self-management activities. This information can be printed out for further reading or added to an individualized *care plan*. The Knowledge Base contains several hundred options for evidence-based self-

management strategies presented in patient-friendly lay language, obtained from scientific literature, validated by an expert panel and is regularly updated. Obviously, the knowledge base is not only helpful to patients, but also to care providers. In *CONNECT* it will be made accessible to them through the HER so that they can use it for evidence-based care planning with their patients.

3. In the *Communication Component* patients can communicate about anything they like. The application's helpdesk system to handle messages from the patients.
 - Data Security. Security and confidentiality are critically important due to the sensitive nature of health information, patient concerns, and institutional requirements. Strong security measures are therefore, implemented into *CONNECT* to protect this information. The current version of *Choice* and *WebChoice* has been approved by the Norwegian Data Inspectorate. In this solution patients are authenticated using a smart card based public key (PKI) solution from Bypass. Support for the BankID PKI solution is also implemented. The confidentiality of the information is protected with 128-bit Secure Sockets Layer (SSL) encryption. Other procedural and technical protections are ensured, such as storing data behind an Internet firewall, robust data integrity and auditing controls, and training of staff who are authorized to access patient data.
 - Support for multiple languages and diagnoses. *CONNECT* will be implemented as a language independent system by separating the content from the system. Language is defined as an attribute of the user.
 - Interfacing *WebChoice* with the EHR. For *CONNECT* we are building a generic integration logic that is built on a service oriented architecture (SOA). Thus *CONNECT* will not be tailored to a specific EHR vendor or just one technological platform. SOA describes how independent services with defined interfaces can be integrated without having previous knowledge of the services that actually perform the tasks. The format of the messages will be based on the EbXML framework as proposed by the Norwegian Centre for Informatics in Health and Social Care (KITH).

Another important goal is that the integration of *Choice/WebChoice* with the EHR only requires minor, if any, changes to the existing EHR system. When a patient has completed a symptom assessment an assessment summary report is generated and a notification message is automatically sent to the EHR. The message contains a Uniform Resource Locator (URL) which is a link to the assessment summary report. The EHR will display this link to the clinicians together with information from other subsystems such as lab or radiology results. When clinicians click on the link they will automatically be redirected to *CONNECT*.

2. *CONNECT* Mobile: Secure, mobile access to the shared EHR

With mobile access to the EHR, we can reach the patient in other situations than only in the hospital and at home. Availability of powerful mobile terminals enable us to

extend access to the shared EHR through a mobile environment. Mobile access to sensitive information systems represents many challenges. To verify and solve critical technical issues, we have chosen a prototype implementation approach, where elements of the *Choice* and *WebChoice* applications are mobilized using standard mobile terminals enhanced with technologies for device-independent display, fonts and other user interface features, as well as mobility and security technologies. The mobile client can detect terminal capabilities and adapt the user interface. We use mobile IP, WLAN/Cellular network detection and Mobile IP for seamless selection of network access. For security, we use 128 bit SSL encryption. The mechanisms for practical authentication on a mobile terminal is still under study. The prototype implementation of *CONNECT* Mobile demonstrates how a context-sensitive, multi-modal patient communication and information tool can be implemented in legacy mobile terminals. A mobile phone, available to most patients, is enhanced using state of the art technologies to ensure adaptation, readability, mobility and security to ensure simple access to the EHR while moving seamlessly through heterogenous networks, and accessing and interacting with the EHR from different locations.

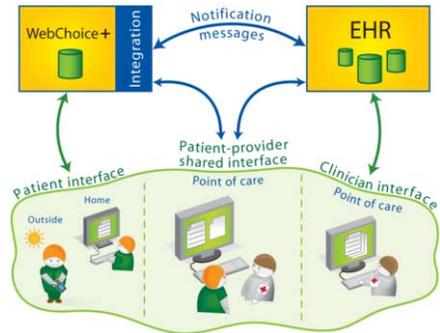


Figure 1: CONNECT component architecture

Key research issues in the project include verification of usability and readability that is needed on a mobile platform with limited display size and diverse user interfaces. We explore how we can create a *WebChoice* client that has the same user interface and ‘look & feel’ on different mobile phones with very diverse technical specifications, and how to create novel methods for secure authentication and encryption of sensitive information in highly mobile environment.

The *CONNECT* mobile client is an extension to the *WebChoice* system. It lets the patient access some of the functionality in *WebChoice* away from home. The mobile client is programmed in Java MIPD 2.0, supported in all recent mobile terminals. The main functionality in the *CONNECT* mobile application includes registration of symptoms, access to database for evidence based information and forum resources. When using the Web-based *WebChoice* the patient needs a computer and must be online. When being away from home it can be difficult to access the system. Using a mobile device we solve the accessibility problem, but are facing other challenges on usability, connectivity, mobility and security challenges addressed in this project.

3. Discussion

While the *CONNECT* application holds great promise to improve patient-provider communication and help patients to better understand and manage their illness, critics may argue that it also could increase the digital divide. While providing all patients with equal access to health services and information are high priority health policy goals in most Western countries, not all patients will be able to benefit from the application. To use *CONNECT*, patients will need a computer, have access to the

Internet or own a cell phone, and are required to have a certain level of computer and health literacy. Thus people who are disadvantaged in the first place may also have the least to gain from the application. It could, however, be argued that helping 70% of the patients is better than helping none. As new technologies mature and more people gain access, these can be better adapted to the needs of people from varying socio-economic backgrounds.

On the clinician side, a number of challenges come into play when developing a new system such as *CONNECT* for routine clinical practice. Resource requirements, feasibility, acceptability and organizational issues need to be carefully addressed. The successful design, implementation and evaluation of clinical support systems depends on how useful clinicians and patients who are the potential main users perceive them. Among possible barriers to clinicians' use of new information systems described in the literature are the organization's attitudes towards innovations, the degree to which the system requires clinicians to modify established routines, and the lack of leadership support.⁶⁻⁸ Clinicians work under time pressure and competing obligations, and additional tasks are not likely to be carried out if not perceived helpful. Many information systems have failed because developers neglected users' judgments about a systems' feasibility, time requirements, and usefulness in clinical practice.^{9,10}

To safe-guard against such problems, *CONNECT* is developed with extensive user-centred design methods. Patients and clinicians participate in all phases of the design process to help us understand their needs for shared decision making, documentation and information sharing; the context they work and live in, their underlying assumptions and expectations and how we best can address them.

On the technical side, a particular challenge is a trade-off between data security, and user friendliness. Because *CONNECT* is to be implemented as a plug-in into existing EHRs that require strict security measures behind safe firewalls, communication with the world outside rises security concerns and challenges related to how advanced authentication mechanisms such as PKI and Smartcards required for access to medical records can be combined with user friendly interfaces that even novel Internet users can master. For example the current *WebChoice* uses a PKI Infrastructure and "My page" technology. While this meets the highest security and privacy standards required by the Norwegian Data Security Act (Datatilsynet) the complexity and number the steps required to install the security measures on their computers have posed problems for a significant number of patients, and also hindered some to get online at all.

Another interface issue relates to compatibility of user interfaces. Every EHR has its own design. Plugging-in *CONNECT* may require different navigation routines and learning needs and not give the user the same touch and feel as the rest of the EHR.

To address these and other challenges the development process of *CONNECT* goes hand in hand with research that includes several doctoral students, to understanding and address these issues and find solutions for how they could potentially be solved.

The clinical systems environment at Rikshospitalet offers a state-of-the-art preview of what many other organizations and systems will offer in the future. By conducting research in this testbed, we can learn both how new interventions are useful, and just as importantly, how to operationalize the interventions. Technology adoption in the area of patient-provider information sharing and collaboration is never automatic. The use of such systems by patients and the willingness by practices to promote their use requires that systems are time neutral or save time and offer conveniences to users. Learning how to bridge the continuum between these approaches is an important area

of learning because of the high potential for benefits to all parties who participate in the care process. The project addresses key technological challenges of information and network security in a mobile environment. Issues like authentication, seamless handover between networks, adaptation to different terminals and secure access to the information stored in the EHR are requirements that must be solved as a part of the project to enable research in a “live” testbed, providing feedback to the interventions and how they can be successfully implemented.

Acknowledgements

This study is supported by two research grants from the Norwegian Research Council and the Norwegian Research Funding program Helse Sor.

References

- 1 Cronin C. Personal Health Records: an overview of what is available to the public. AARP Public Policy Institute. 2006.
- 2 Connecting for health: a public-private collaborative: final report. The Personal health working group. Markle Foundation, 2003.
- 3 Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. *Jamia* 2006;13: 121-26.
- 4 Ruland, C.M., Decision support for patient preference-based care planning: effects on nursing care and patient outcomes. *J Am Med Inform Assoc*, 1999. 6(4): p. 304-12.
- 5 Ruland, C.M., Handheld Technology to Improve Patient Care: Evaluating a Support System for Preference-based Care Planning at the Bedside. *J.Am.Med.Inform.Assoc.*, 2002. 9(2): p. 192-201.
- 6 Schultz, R.L. and D.P. Slevin, Implementation and organizational Validity: An empirical investigation, in *Implementing Operations Research/Management Science*, R.L. Schultz and D.P. Slevin, Editors. 1975, Elsevier: New York, NY. p. 153-182.
- 7 Rogers, E.M., *Diffusion of innovations*. 3rd ed. 1983, New York, NY: The Free Press.
- 8 Davis, F.D., Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quarterly*, 1989(13): p. 319-340.
- 9 Anderson, J., G., C. Aydin, and S.J. Jay, eds. *Evaluating health care information systems. Methods and applications*. 1994, SAGE publications: Thousand Oaks.
- 10 Holmes-Rovner, M., et al., Implementing shared decision-making in routine practice: barriers and opportunities. *Health Expect*, 2000. 3(3): p. 182-191.
- 11 Patel, V. and A.W. Kushniruk, Human computer interaction in health care, in *Handbook of medical informatics*, v.J.H. Bommel and M.A. Musen, Editors. 1997, Springer: Diegem. p. 473-493.

Examination of computer assisted prescribing of an initial calculated antibiotic treatment

RÖHRIG R^a, NICZKO EJ^b, BEUTEFÜHR H^c, BÖTTGER S^d,
KLASEN J^e, FÜSSLE R^f, HARTMANN B^{e,1}

^a*Department of Anaesthesiology, Intensive Care Medicine, and Pain Therapy
Justus-Liebig-University of Giessen, Germany*

^b*Department of Anaesthesiology and Intensive Care Medicine, University of Magdeburg*

^c*Department of Neurology and Psychiatry, Clinical Centre Braunschweig*

^d*Department of Oral and Maxillofacial Surgery, University of Magdeburg*

^e*Department of Anaesthesiology and Intensive Care Medicine,
Hospitals of Ostallgaeu-Kaufbeuren*

^f*Department of Medical Microbiology and Virology,
Justus-Liebig-University of Giessen*

Abstract: The objective of this prospective clinical usage study was to examine the value of the rule based ‘Therapeutic Assistant’ integrated into an existing Patient Data Management System (PDMS) in helping to prescribe a initial antibiotic regime in accordance with the requirements of accepted guidelines. A prospective study comparing data before and after the introduction of the ‘Therapeutic Assistant’ was carried out. An adequate therapy resulted significantly more often after the introduction of the ‘Therapeutic Assistant’ [$p < 0.05$]; however no difference between the regimes with and without the ‘Therapeutic Assistant’ in the period after its introduction could be established. Whether the ‘Therapeutic Assistant’ influenced the prescriptions made without it will have to be established in a further study.

Keywords: Protocol Drug Therapy, Computer-Assisted, Empirical Antimicrobial Therapy, Patient Data Management System

1. Introduction

In intensive care patients the occurrence of infections and their prompt and adequate treatment is a major factor in determining the eventual outcome for the patient. Apart from the surgical elimination of the source of the infection, the initial calculated antibiotic regime decisively influences the further course of the infection (10-12).

There are various strategies available to improve the initial calculated antibiotic regime (9;16). The introduction of information technology directly at the patient’s bedside provides a possible alternative in the form of systems which support in

¹ Address for correspondences: E-Mail: Bernd.Hartmann@klinikum-kf-oal.de

decision making. (2;8;12;18). In the development of these systems it was repeatedly shown that, under laboratory conditions, they achieved their required goals. However they often fail in their routine use (19;12).

The objective of this clinical usage study was to examine the value of the rule based ‘Therapeutic Assistant’ integrated into an existing Patient Data Management System (PDMS) in helping to prescribe a calculated antibiotic regime in accordance with the requirements of accepted guidelines. In particular, changes in the quality of the treatment should be critically considered.

2. Study Context

2.1. System description

Since 1999 the surgical Intensive Care Unit (SICU) of the University Hospital, Giessen uses a PDMS (ICUData, Imeso GmbH) to document all medical data, including physicians’ orders (14). The PDMS includes a decision support system for therapeutic guideline assistance since version 3.0 to help maintain standards when writing orders. As they are only suggestions, they are initially recorded in the chart as “not cleared” and must then be manually ordered by the physician. At this point, the physician can make any changes to the suggested therapy as needed. As such, the responsibility for therapeutic decisions remains with the physician.

Further co-operation with our Centre for Medical Microbiology and Virology has allowed to integrate an ‘Antibiotic Wizard’ module within the electronic patient record. This contains a catalogue of rule-based therapy standards for empirical antimicrobial therapy and is representative of the therapy recommendations of the Paul Ehrlich Gesellschaft (PEG) (21), modified for ICU in the spectra of resistances and pathogens. The graphical user interface (GUI) leads the physician to create a treatment plan based on rules defined in the database and uses a question and answer format (17).

The “Antibiotic Wizard” is designed to assist the ward physician in empirical antimicrobial therapy. Based on the PEG guidelines, the physician is given the standard operating procedures (SOP) available for treatment. This is meant to improve the quality of orders written, in that the program helps standardise antibiotic doses, combinations and duration of treatment. There is also a reminder feature that is integrated in the patient’s chart, which after two or three days, shows the physician the antibiotic therapy ordered. This is to ensure that the antibiotic therapy is reassessed for the continuing presence of pathogens and to make any adjustments necessary (17).

2.2. Organisational Setting

The survey was done in the SICU with 14 beds and a throughput of approximately 1500 patients annually. Twice weekly (Tuesday and Friday) there is a microbiological (ward) round with a microbiologist together with the SICU Registrar (senior resident) and the house officers (residents). The current microbiological findings are discussed in these rounds, together with the antibiotic treatment as well as the patients’ infective problems. Further therapeutic ideas are considered and ordered.

3. Methods

3.1. Study design

The microbiological rounds were accompanied regularly for six months before the planned training and introduction of the ‘Therapeutic Assistant’ “antibiotic Wizard” and this was continued after the introduction as part of the clinical routine (pre- and post-periods.) The survey covered a time span of 14 months, of which the pre-period was 142 days and the post-period was 272 days.

The physicians were asked about the indications of the calculated antibiotic regimes. They had to judge whether this therapy was successful or if there were reasons for an alteration of therapy. The standardised information for the study was entered in the PDMS in the course of the rounds. Only patients who had had at least one standardised round visit documented were included in the pre- and post-periods comparison. Patients who were treated in ICU during the time between two microbiological rounds could not be taken into account.

All the data necessary for the study were extracted from the PDMS using SQL scripts. These included among others age, sex, length of stay in ICU, duration of ventilation and deaths in ICU. The SAPS II score (13), the daily SOFA score (20) and the delta SOFA (15) were collected automatically (3;7).

3.2. Measurement of Outcome

On foot of the documented microbiological rounds and considering the microbiological findings a blind judgement was made — the antibiotic therapy was “adequate” or “inadequate” for its main purpose. Further outcome parameters were ICU mortality, the length of stay in intensive care, the duration of ventilation and the delta-SOFA, which counts as a measure of the decline in organ function whilst in the ICU and is therefore potentially influenced by therapeutic measures.

3.3. Data analysis

Using the Mann-Whitney-U test or the Chi² test significant differences ($p < 0.05$) between the two observation periods were sought. On the one hand, the period before the introduction of the “Antibiotic Wizard” (pre-period) was compared with the period afterwards (post-period); on the other hand a sub-group analysis was performed on the patients who had had at least one antibiotic regime proposed by the “Antibiotic Wizard” and the whole patient group in both the pre- and post-periods with the exception of the “Antibiotic Wizard” sub group. Three comparative group analyses in total were carried out to verify any possible influence of the rule-based ‘Therapeutic Assistant’ on the “adequate calculated antibiotic regime” in the quality of patient treatment. A further multivariable analysis to examine possible degree of influence on the outcome parameters would have been meaningless due to the limited number of cases.

4. Results

There were 43 patients in the pre-period who had documented microbiological rounds; the “Antibiotic Wizard” was used in 35 out of the 113 post-period patients. Table 1 shows the whole patient group; table 2 shows the statistics for the comparative group analyses.

Table 1: All patients and Outcome parameters (bold)

MV = Mean Value, SD = Standard deviation, CI = Confidence interval, IQS = Interquartile range

parameters	Pre-Period (43 Patients)				Post-Period (113 Patients)			
	MV ± SD	CI	Median	IQR	MV ± SD	CI	Median	IQR
age [Years]	59,9 ± 20,1	[53,7; 66,1]	66,1	[45,1; 76,2]	56,9 ± 18,9	[53,4; 60,4]	61,3	[44,3; 71,0]
SAPS II	49 ± 15	[44,5; 53,8]	44	[40; 61]	48 ± 12	[45,4; 49,8]	48	[40; 54]
SOFA (Admission at ICU)	6,4 ± 2,6	[5,6; 7,2]	6	[5; 8]	6,1 ± 2,4	[5,7; 6,6]	6	[4; 8]
Delta-SOFA	1,9 ± 2,3	[1,2; 2,7]	1	[0; 3]	1,4 ± 1,9	[1,0; 1,8]	1	[0; 2]
length of stay (ICU) [h]	472 ± 538	[307; 638]	376	[209; 524]	337 ± 275	[286; 389]	270	[122; 479]
duration of ventilation [h]	254 ± 274	[169; 338]	197	[46; 365]	178 ± 207	[139; 217]	106	[14; 276]
adequate therapy	47,8 ± 30,7	[38,3; 57,2]	50	[33,3; 68,2]	66,5 ± 30,5	[60,8; 72,2]	75	[50; 91,3]
evaluation of therapy success	34,3 ± 43,0	[21,0; 47,5]	0	[0; 100]	63,6 ± 44,2	[55,4; 71,9]	100	[0; 100]
	yes	no	%		yes	no	%	
gender (male)	32	11	74,4		67	46	59,3	
mortality in ICU	18	25	41,9		30	83	26,5	
parameters	Post-Period ordered by Wizard (35 Patients)				Post-Period without Wizard (78 Patients)			
	MV ± SD	CI	Median	IQR	MV ± SD	CI	Median	IQR
age [Years]	57,2 ± 20,4	[50,2; 64,2]	66,7	[43,34; 72,93]	56,8 ± 18,3	[52,7; 60,9]	59,2	[45,6; 69,6]
SAPS II	45 ± 12	[40,8; 49,3]	43	[35; 53]	49 ± 11	[46,2; 51,3]	49	[42; 55]
SOFA (Admission at ICU)	6,2 ± 2,5	[5,3; 7,0]	5	[4; 8]	6,1 ± 2,4	[5,5; 6,6]	6	[4; 8]
Delta-SOFA	1,1 ± 1,4	[0,6; 1,6]	1	[0; 2]	1,6 ± 2,1	[1,1; 2,0]	1	[0; 3]
length of stay (ICU) [h]	388 ± 300	[285; 491]	296	[136; 645]	315 ± 262	[255; 374]	234	[120; 445]
duration of ventilation [h]	189 ± 249	[104; 275]	99	[10; 216]	173 ± 187	[130; 215]	108	[17; 284]
adequate therapy	72,5 ± 26,3	[63,4; 81,5]	75	[53,1; 100]	63,8 ± 32,1	[56,5; 71,0]	74	[50; 87,9]
evaluation of therapy success	64,8 ± 41,4	[50,6; 79,0]	100	[33,3; 100]	63,1 ± 45,7	[52,8; 73,4]	100	[0; 100]
	yes	no	%		yes	no	%	
gender (male)	22	13	62,9		45	33	57,7	
mortality in ICU	6	29	17,1		24	54	30,8	

Table 2: Statistics for the comparative group analyses

parameters	Mann-Whitney-U-Test (p-Value)		
	pre-period vs. post-period	pre-period vs. Wizard-Pat.	Wizard-Pat. vs. Post (Ø Wizard)
age [Years]	0,28	0,52	0,55
SAPS II	0,94	0,29	0,08
SOFA (Admission at ICU)	0,52	0,58	0,98
Delta-SOFA	0,23	0,18	0,63
length of stay (ICU) [h]	0,07	0,51	0,29
duration of ventilation [h]	0,07	0,13	0,71
adequate therapy	< 0,01	< 0,01	0,24
evaluation of therapy success	< 0,01	< 0,01	0,97
	X ² -Test (p-Value)		
gender (male)	0,10	0,33	0,68
mortality in ICU	0,06	< 0,05	0,17

5. Discussion

It was shown that, in the post-period in which the “Antibiotic Wizard” was available, the assessment of the antibiotic therapy as adequate was significantly more frequent and less frequently assessed as inadequate, and the treatment was more often evaluated as successful. In the patient group whose initial antibiotic regime was prescribed using the “Antibiotic Wizard” the relative frequency of adequate treatment from an average of 47.8% in the pre-period to 72.5% in the post-period. At the same time the rate of inadequate treatment decreased from 34.2% to 18.5% in average. In a large multicentre trial of 904 patients Harbarth et al. showed a similarly high incidence of inadequate initial antibiotic treatment in 23% of cases (5). The total mortality in this study was 24% (168/693) among the adequately treated cohort compared to 39% (82/211) in those with inadequate initial therapy (5). Ibrahim and Leibovici demonstrated that an initially inadequate antibiotic treatment clearly increased the mortality (61.9% vs. 28.4% or 34.3% vs. 20.2% (6)). In spite of a smaller case number in this study and the absence of proof of a correlation between adequate therapy and ICU mortality, a similar reduction in mortality from 40.9% in the pre-period to 26.5% in the post-period was observed. In those patients whose antibiotic regimes was prescribed using the ‘Therapeutic Assistant’ the mortality rate was 17.1%, which was significantly less than in the pre-period.

The treatment improved in the post-period both with and without the wizard. How far this is a advantage of the “Antibiotic Wizard” needs to be examined. Besides the use of the ‘Therapeutic Assistant’ as a reference work without using it to prescribe, the documentation of the microbiological rounds may have led to an increased level of care in the prescribing of antibiotic treatment by the physicians. It is therefore possible that another study might show that, in spite of computer assisted prescribing, the communication between ICU physicians and infection specialists is responsible for the correct antibiotic usage (1).

The reasons for the limited prescribing using the “Antibiotic Wizard” need to be ascertained. Other studies have shown that, through the use of a decision support system, the prescription of an empirical initially adequate antibiotic regime rose from an initial 77% to 94% (4). In our study this high improvement rate could not be achieved. However the conclusions, with an improvement of initial antibiotic treatment adjudged to be adequate from 47.8% to 72.5%, are very promising.

6. Acknowledgements

We would like to thank MoReData GmbH in Giessen, Germany for their help in statistical analysis and Dr. Brendan Whelan (Dublin, Ireland) for reviewing the manuscript.

7. References

- (1) Alvarez-Lerma F, Palomar M, Grau S. Management of antimicrobial use in the intensive care unit. *Drugs* 2001;61(6):763-75.
- (2) Delpierre C, Cuzin L, Fillaux J, Alvarez M, Massip P, Lang T. A systematic review of computer-based patient record systems and quality of care: more randomized clinical trials or a broader approach? *Int J Qual Health Care* 2004 Oct;16(5):407-16.

- (3) Engel, J., Junger, A., Böttger, .. et al. Outcome prediction in a surgical ICU using automatically calculated SAPS II scores. *Anaesth Intensive Care* 2003 Oct;31(5):548-54.
- (4) Evans RS, Classen DC, Pestotnik SL, Lundsgaarde HP, Burke JP. Improving empiric antibiotic selection using computer decision support. *Arch Intern Med* 1994 Apr 25;154(8):878-84.
- (5) Harbarth S, Garbino J, Pugin J, Romand JA, Lew D, Pittet D. Inappropriate initial antimicrobial therapy and its effect on survival in a clinical trial of immunomodulating therapy for severe sepsis. *Am J Med* 2003 Nov;115(7):529-35.
- (6) Ibrahim EH, Sherman G, Ward S, Fraser VJ, Kollef MH. The influence of inadequate antimicrobial treatment of bloodstream infections on patient outcomes in the ICU setting. *Chest* 2000 Jul;118(1):146-55.
- (7) Junger A, Böttger S, Engel J, Benson M, Michel A, Röhrig R, et al. Automatic calculation of a modified APACHE II score using a patient data management system (PDMS). *Int J Med Inform* 2002 Jun;65(2):145-57.
- (8) Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ* 2005 Apr 2;330(7494):765.
- (9) Kollef M. Appropriate empirical antibacterial therapy for nosocomial infections: getting it right the first time. *Drugs* 2003;63(20):2157-68.
- (10) Kollef MH. The importance of appropriate initial antibiotic therapy for hospital-acquired infections. *Am J Med* 2003 Nov;115(7):582-4.
- (11) Kollef MH, Ward S, Sherman G, Prentice D, Schaiff R, Huey W, et al. Inadequate treatment of nosocomial infections is associated with certain empiric antibiotic choices. *Crit Care Med* 2000 Oct;28(10):3456-64.
- (12) Kunz JC, Shortliffe EH, Buchanan BG, Feigenbaum EA. Computer-assisted decision making in medicine. *J Med Philos* 1984 May;9(2):135-60.
- (13) Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993 Dec 22;270(24):2957-63.
- (14) Michel A, Benson M, Junger A, Sciuk G, Hempelmann G, Dudeck J, et al. Design principles of a clinical information system for intensive care units (ICUData). *Stud Health Technol Inform* 2000;77:921-4.
- (15) Moreno R, Vincent JL, Matos R, Mendonca A, Cantraine F, Thijs L, et al. The use of maximum SOFA score to quantify organ dysfunction/failure in intensive care. Results of a prospective, multicentre study. Working Group on Sepsis related Problems of the ESICM. *Intensive Care Med* 1999 Jul;25(7):686-96.
- (16) Niederman MS. Appropriate use of antimicrobial agents: challenges and strategies for improvement. *Crit Care Med* 2003 Feb;31(2):608-16.
- (17) Röhrig R, Beuteführ H, Hartmann B, Niczko E, Quinzio B, Junger A, et al. Summative Software Evaluation of a Therapeutic Guideline Assistance System for Empiric Antimicrobial Therapy in ICU. *J Clin Monit Comput* 2007 Apr 4.
- (18) Shortliffe EH, Axline SG, Buchanan BG, Merigan TC, Cohen SN. An artificial intelligence program to advise physicians regarding antimicrobial therapy. *Comput Biomed Res* 1973 Dec;6(6):544-60.
- (19) Shortliffe EH, Davis R, Axline SG, Buchanan BG, Green CC, Cohen SN. Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the MYCIN system. *Comput Biomed Res* 1975 Aug;8(4):303-20.
- (20) Vincent JL, Moreno R, Takala J, Willatts S, De MA, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996 Jul;22(7):707-10.
- (21) Vogel F, Bodmann K. Empfehlungen zur kalkulierten parenteralen Initialtherapie bakterieller Erkrankungen bei Erwachsenen. *Chemotherapie Journal* 2004;13:46-105.
- (22) Yu VL, Buchanan BG, Shortliffe EH, Wraith SM, Davis R, Scott AC, et al. Evaluating the performance of a computer-based consultant. *Comput Programs Biomed* 1979 Jan;9(1):95-102.

2. Consumer and Home Based eHealth

This page intentionally left blank

Treasure Hunt – a serious game to support psychotherapeutic treatment of children

Veronika BREZINKA¹

Department of Child and Adolescent Psychiatry, Zürich University, Zürich, Switzerland

Abstract. Computer and video games for children have gained negative publicity due to reported associations between intensive gaming and aggressive behaviour, school failure, and overweight. While most studies centre upon negative consequences of video games, their innovative potentials tend to be overlooked. One field for the innovative use of computer games is child psychotherapy. By including therapeutic concepts into a video game, children can be offered attractive electronic homework assignments that enable them to rehearse and repeat basic psychoeducational concepts they have learned during therapy sessions. Moreover, therapeutic games can help therapists to structure therapy sessions. Psychotherapeutic computer games translated into foreign languages could form a useful tool in the treatment of migrant children. ‘Treasure Hunt’ is the first serious game based on principles of cognitive behaviour modification. It is developed for eight to twelve year old children who are in cognitive-behavioural treatment for various disorders. Reactions of children and therapists to experimental versions of the game are positive. Serious games might prove a useful tool to support psychotherapeutic treatment of children.

Keywords. Home-based e-Health, Human-Computer interaction, learning and education

Introduction

Although computers and internet are a normal part of life for millions of children, they have gained mainly negative publicity due to the reported association between intensive gaming and aggressive behaviour, game addiction, school failure and overweight [1-3]. As a consequence, most reports on the effects of video games centre upon their potential negative consequences [4-6], while their innovative potentials tend to be overlooked [7].

¹ Adress for correspondences: Universität Zürich, Child and Adolescent Psychiatry
Neptunstr. 60, CH-8032 Zürich. E-mail: veronika.brezinka@ppkj.uzh.ch

However, these innovative potentials exist. Computer games improve spatial performance in children, adolescents and adults [8]. Action-video-game playing with so called first person shooters has been reported to alter a range of visual skills and to enhance visuospatial attention [9, 10]. Studies also show that commercial computer games can be used innovatively at school [11]. In the medical sector, various reports show that computer games have been therapeutically successful with physically handicapped children and children with chronic disease like asthma [12], diabetes [13] and cancer [14, 15]. As for psychotherapy, it would be difficult for child therapists to ignore how fascinated children are by computers and video games. Yet, up to now few initiatives make use of this fascination to enhance psychotherapeutic treatment. There are two exceptions: the game ‘Earthquake in Ziiland’, based on family therapy, is designed to support psychotherapy of children whose parents have divorced (www.ziilandinteractive.com). ‘Personal Investigator’, a therapeutic game based on solution focused therapy, tries to motivate adolescents for psychotherapy [16].

Developing a computer-based treatment for children does not simply mean putting a treatment manual on the screen. In order to motivate children for psychotherapy, the challenge lies in designing serious games that incorporate therapeutic goals as well as various skills training. By including therapeutic concepts into a game, children can be offered attractive electronic homework assignments that enable them to repeat and rehearse basic psychoeducational concepts they have learned during therapy sessions. Moreover, a psychotherapeutic game that matches the theoretical orientation of the therapist can help him / her to structure therapy sessions and to explain important theoretical concepts in diverse ways and with a metaphor and means attractive to most children.

1. Method

Cognitive behaviour therapy is one of the best-researched and empirically supported treatment methods for adults and children. Its theoretical framework is based on the assumption that emotions and behaviour are largely a product of cognitions; thus psychological and behaviour problems can be reduced by altering cognitive processes. Cognitive behaviour therapy offers several intervention programmes for children from which learning goals for serious games can be derived.

In order to support cognitive-behavioural treatment of children, the Department of Child and Adolescent Psychiatry of Zürich University is developing a serious game called ‘Treasure Hunt’. The game is not meant to substitute the therapist, but to support therapy by offering electronic homework assignments and rehearsing basic psychoeducational parts of treatment. The theoretical background of the game is formed by basic concepts of cognitive behaviour therapy for children as outlined in treatment programmes like ‘Coping Cat’ [17], ‘Friends for Children’ [18] and ‘Keeping your cool’ [19]. These basic concepts are important for the treatment of both internalizing and externalizing disorders.

Treasure Hunt is an interactive adventure game with six levels. Each of the six levels of the game corresponds to a certain step in cognitive-behavioural treatment. The maximum amount of time needed to solve all tasks of a level is about twenty minutes.

1.1. Story

The game takes place aboard an old ship inhabited by Captain Jones, Felix the ship's cat and Polly the ship's parrot. Captain Jones has found an old treasure map in the hull of his ship. To solve its mystery, he needs the help of a child. Tasks take place in different parts of the ship – on deck, in the galley, in the dining room of Captain Jones and in the shipmates' bunks. Each task corresponds to a certain step in cognitive-behavioural treatment, implying a linear structure of the game. For each completed task, the child receives a sea star. The old treasure map has a dark spot in the shape of a sea star at six important places. The child and Captain Jones will only be able to read what is written there when they place the missing sea star on the map. After having solved all the tasks, the last mission consists of a recapitulation of the previous exercises. Once the child has solved this last problem, he/she will find out where the treasure is buried. One of the most interesting parts of the game is dedicated to the hunting of unhelpful (automatic) thoughts by means of an ego-shooter. The child has to catch a flying fish to be able to read the unhelpful thought written on it and replace it by a helpful one.

Before joining Captain Jones on the final search for the treasure, the child receives a sailor's certificate that summarizes what he/she has learnt through the game and that is signed by Captain Jones and the therapist.

1.2. Software and Tools

Treasure Hunt is a 2.5 D Flash adventure game programmed with Actionscript and XML. Flash was used to guarantee platform independence, as only a Flash compatible internet browser is needed and no programme has to be installed. This facilitates giving homework to children independent from their computer hardware and operating system at home. User interaction will be recorded in XML files to help therapists analyze children's choices and / or progress.

2. Results

Playability tests with an experimental version showed that children appreciate the game and its diverse tasks. Several therapists in the Department of Child and Adolescent Psychiatry of Zürich University have used pilot versions of the game with children treated for a variety of disorders (anxiety disorder, depression, behavioural disorder). Because they worked with an experimental version, therapists were not allowed to give parts of the game as homework assignment, but instead used it *during* therapy sessions. They reported to use the game as reinforcement, telling the child 'if you work well, we will play Treasure Hunt for the last ten minutes'. All therapists reported positive reactions of the children in treatment and liked to work with the game themselves. Although Treasure Hunt was originally developed to offer attractive homework assignments in between therapy sessions, the pilot showed that the game can also help young or less experienced therapists to structure therapy sessions and to explain important cognitive-behavioural concepts like the influence of thoughts on our feelings or the distinction between helpful and unhelpful thoughts.

While these primary findings are encouraging, conclusions about the effectiveness of Treasure Hunt are premature, as the professional version of the game is not finished

yet. Treasure Hunt is not developed as a self-help instrument and should be used under guidance of a behaviour therapist. Our pilot showed that using a computer game in psychotherapy sessions does not mean that classic therapeutic methods like writing, drawing or role-playing lose their significance in the treatment of children and adolescents – on the contrary, as various exercises for further therapy sessions can be derived from Treasure Hunt. Therapists reported for example that they asked children to ‘help us to design a next level’ or to ‘draw flying fish with more unhelpful thoughts’.

Whether Treasure Hunt can really support cognitive-behavioural treatment of children and enhance child compliance for homework assignments will be evaluated as soon as the professional version of the game is finished.

3. Discussion and Conclusion

Psychotherapy of children and adolescents is an area in which innovative use of computers in the form of psychotherapeutic video games may enhance child compliance and offer new ways of treatment. Such games have the potential to enhance child compliance, offer attractive homework assignments, structure therapy sessions and support treatment of migrant children who could play the games in their own language and share their content with parents and siblings.

Cognitive behaviour therapy offers several intervention programmes for children from which learning goals for therapeutic video games can be derived. For example social problem solving, a standard therapeutic intervention for young children [20, 21] could be incorporated into a game to support psychotherapy with children as young as five years. Such a game might even be used for the prevention of behaviour problems in this age group. Incorporating elements of anger management programmes [19, 22] into video games could be an even greater challenge for the development of serious games. As most of the children treated for anger and aggression problems are boys, and boys are reported to show considerably more fascination for computers than girls [8], creating serious games that include anger management strategies might support treatment of this notoriously difficult and non compliant group. An alternative pathway could be the development of therapeutic games built on Dodge’s theory of social-cognitive biases of aggressive children [23]. If such games could help aggressive children to reduce hostile attributional biases and to ameliorate cognitive processing of potentially threatening situations [24, 25], treatment of a chronic and difficult group of clients might become easier.

There is, however, still a long way to go and considerable resistance to overcome. Not all game-designers are positive about the concept of serious games, suggesting that a game that *has to be played* might lose its attractiveness. On the other hand, many academics and health professionals are not used to view computer games as something different from ‘pure fun’ or ‘only a game’ and doubt that a computer game can teach useful skills. Moreover, there is fear that if psychotherapeutic games are successful, computers might replace therapists in the long run. However, no psychotherapeutic game will be able to alleviate childhood problems on its own, and therapeutic games will show their maximum potentials only under guidance of a therapist who can explain and comment the concepts introduced in the game.

New developments in gaming technology as well as more research on therapeutic games will hopefully lead to the creation of more serious games to support child

psychotherapy. Ideally, these games should be labelled with a quality seal for therapists.

Acknowledgments

I thank the Department of Child and Adolescent Psychiatry (Head: Prof. Dr. Dr. H.C. Steinhausen) and the Technology Transfer Fund of Zürich University for financial support. Special thanks are due to Markus Ruh (programming), Christoph Wartmann (design), Denis Chait (renderings) and Wolfgang Güdden (music).

Bibliography

- [1] Browne, K.D. and C. Hamilton-Giachritsis, *The influence of violent media on children and adolescents: a public-health approach*. *Lancet*, 2005. **365**(9460): p. 702-10.
- [2] Huesmann, L.R., J. Moise-Titus, C.L. Podolski, et al., *Longitudinal relations between children's exposure to TV violence and their aggressive and violent behavior in young adulthood: 1977-1992*. *Developmental Psychology*, 2003. **39**(2): p. 201-21.
- [3] Slater, M.D., K.L. Henry, R.C. Swaim, et al., *Violent media content and aggressiveness in adolescents: A downward spiral model*. *Communication Research*, 2003. **30**(6): p. 713-736.
- [4] Carnagey, N.L. and C.A. Anderson, *The effects of reward and punishment in violent video games on aggressive affect, cognition, and behavior*. *Psychological Science*, 2005. **16**: p. 882-889.
- [5] Funk, J.B., *Children's exposure to violent video games and desensitization to violence*. *Child & Adolescent Psychiatric Clinics of North America*, 2005. **14**: p. 387-404.
- [6] Gentile, D.A., P.J. Lynch, J.R. Linder, et al., *The effects of violent video game habits on adolescent hostility, aggressive behaviors, and school performance*. *Journal of Adolescence*, 2004. **27**(1): p. 5-22.
- [7] Griffiths, M., *The therapeutic use of videogames in childhood and adolescence*. *Clinical Child Psychology and Psychiatry*, 2003. **8**(4): p. 547-554.
- [8] Subrahmanyam, K., P. Greenfield, R. Kraut, et al., *The impact of computer use on children's and adolescents' development*. *Applied Developmental Psychology*, 2001. **22**: p. 7-30.
- [9] Green, C.S. and D. Bavelier, *Effect of Action Video Games on the Spatial Distribution of Visuospatial Attention*. *Journal of Experimental Psychology: Human Perception & Performance*, 2006. **32**(6): p. 1465-1478.
- [10] Green, C.S. and D. Bavelier, *Action video game modifies visual selective attention*. *Nature*, 2003. **423**: p. 534-537.
- [11] Jayakanthan, R., *Application of computer games in the field of education*. *The Electronic Library*, 2002. **20**(2): p. 98-102.
- [12] Lieberman, D.A., *Management of chronic pediatric diseases with interactive health games: Theory and research findings*. *Journal of Ambulatory Care Management*, 2001. **24**(1): p. 26-38.
- [13] Brown, S.J., D.A. Lieberman, B.A. Gemeny, et al., *Educational video game for juvenile diabetes: Results of a controlled trial*. *Medical Informatics*, 1997. **22**(1): p. 77-89.
- [14] Redd, W.H., P.B. Jacobsen, I.M. Die Tril, et al., *Cognitive-attentional distraction in the control of conditioned nausea in pediatric cancer patients receiving chemotherapy*. *Journal of Consulting & Clinical Psychology*, 1987. **55**: p. 391-395.
- [15] Vasterling, J., R.A. Jenkins, D.M. Tope, et al., *Cognitive distraction and relaxation training for the control of side effects due to cancer chemotherapy*. *Journal of Behavioral Medicine*, 1993. **16**: p. 65-80.
- [16] Coyle, D., M. Matthews, J. Sharry, et al., *Personal Investigator: A therapeutic 3D game for adolescent psychotherapy*. *Journal of Interactive Technology & Smart Education*, 2005. **2**: p. 73-88.
- [17] Kendall, P.C., *Coping Cat Workbook*. 1990: Available from P.C. Kendall, Department of Psychology, Temple University, Philadelphia, PA 19122.
- [18] Barrett, P., H. Lowry-Webster, and C. Turner, *Friends for Children Workbook*. 2000, Bowen Hills: Australian Academic Press.
- [19] Nelson, W.M. and A.J. Finch, *'Keeping Your Cool': Cognitive-behavioral therapy for aggressive children: Therapist manual*. 1996, Ardmore, PA: Workbook Publishing.

- [20] Shure, M.B. and G. Spivack, *Interpersonal problem solving as a mediator of behavioral adjustment in preschool and kindergarten children*. Journal of Applied Developmental Psychology, 1980. **1**(1): p. 29-44.
- [21] Webster-Stratton, C., M.J. Reid, and M. Hammond, *Treating Children With Early-Onset Conduct Problems: Intervention Outcomes for Parent, Child, and Teacher Training*. Journal of Clinical Child & Adolescent Psychology, 2004. **33**(1): p. 105-124.
- [22] Lochman, J.E. and L.A. Lenhart, *Anger coping intervention for aggressive children: Conceptual models and outcome effects*. Clinical Psychology Review, 1993. **13**(8): p. 785-805.
- [23] Dodge, K.A., *Translational science in action: Hostile attributional style and the development of aggressive behavior problems*. Development and Psychopathology, 2006. **18**: p. 791-814.
- [24] Dodge, K.A., J.M. Price, J.A. Bachorowski, et al., *Hostile Attributional Biases in Severely Aggressive Adolescents*. Journal of Abnormal Psychology, 1990. **99**(4): p. 385-392.
- [25] Dodge, K.A., J.E. Lochman, J.D. Harnish, et al., *Reactive and proactive aggression in school children and psychiatrically impaired chronically assaultive youth*. Journal of Abnormal Psychology, 1997. **106**(1): p. 37-51.

Virtual Rehabilitation after Stroke

Jurgen BROEREN^{a, b, c}, Ann BJORKDAHL^{b, c}, Lisbeth CLAEISSON^{b, c}, Daniel GOUDE^a, Åsa LUNDGREN-NILSSON^{b, c}, Hans SAMUELSSON^b, Christian BLOMSTRAND^b, Katharina S SUNNERHAGEN^b, Martin RYDMARK^a

^a *Mednet, Institute of Biomedicine, Göteborg University, Göteborg, Sweden*

^b *Institute of Neuroscience and Physiology, Göteborg University, Göteborg, Sweden*

^c *Department of Occupational therapy, Sahlgrenska University hospital, Göteborg, Sweden*

Abstract. The purpose of this project was to investigate the effects of Virtual Reality technology and haptics for stroke rehabilitation. Twenty-nine stroke subjects, 17 women, and 12 men aged 44-85 years, participated in three different studies. All participants responded favorable to the use of the VR activity station. A change of attitude took place after the subjects were exposed to playing computer games. The general experience with the VR application approach suggests that this treatment concept is promising in stroke rehabilitation, with a wide range of applicability.

Key words. haptics, rehabilitation, stroke, virtual reality

Introduction

Stroke is one of our most widespread diseases in the Western world and the principal cause of permanent physical impairment in the adult population in Sweden [1]. One of the most striking social consequences of stroke is a failure to resume activities, which are purely for enjoyment [2]. In addition, enjoyable and pleasurable activities seem to be even further decreased in older adults who have experienced a stroke [3]. This striking decrease is said to be due to depression, upper extremity motor dysfunction, or decreased visuo-spatial ability [4].

In the last decade of the 20th-century, virtual reality technologies first began to be developed and studied as potential tools for assessment and treatment in rehabilitation [5]. The idea behind creating Virtual Reality (VR) is simple; a simulated world runs on a computer system. Training with haptic devices has been suggested to enhance stroke rehabilitation using VR [6-8].

Employing computer games to enhance training motivation is an opportunity, illustrated by the growing interest in the field of Serious Games (www.seriousgames.org). A serious game is a computer-based game with the goal of education and/or training in any form. This stands in contrast to traditional computer games, whose main purpose is to entertain. Serious games include games for learning, games for health, and games for policy and social change. The healthcare sector is showing a steadily increasing interest in serious games. Integrating gaming features

into virtual environments has been reported to enhance motivation in adults undergoing physical and occupational therapy following a stroke [9].

The aim of this project was twofold: (1) to assess the application of a VR activity station as an occupational therapy assessment/treatment method in a hospital and non-hospital environment to see if it could be used there and to evaluate whether playing 3D computer games resulted in improved motor function in persons with prior stroke. (2) to apply the VR activity station as a tool for assessing visuospatial neglect.

1. Material and methods

1.1. Subjects

Twenty-nine stroke subjects, 17 women and 12 men, aged 44-85 years, post stroke 1-140 months, participated in three different studies. Thirteen subjects underwent in- or out-patient rehabilitation at the Department of Rehabilitation Medicine at Sahlgrenska University Hospital, and 16 underwent rehabilitation in a facility for community dwelling persons with stroke (Stroke Forum) in Göteborg, Sweden. A group of healthy individuals (self-perceived health) were recruited and served as reference subjects (range 33-85 years). The study was approved by the Ethics Committee of Göteborg University (Ö511-01), and all participants gave their informed consent.

1.2. System

A VR activity station was used (figure 1). The user stood in the real world and looked into a virtual world generated in the computer. He or she was now able to reach into a virtual space and interact with 3-dimensional (3D) objects through a handheld stylus (haptic device) positioned in the line of sight. It created an illusion of virtual objects for the user while the only real element was the handheld stylus and the computer equipment. Using stereoscopic shuttered glasses, the user observed a 3D image displayed above the tabletop. Every time an activity was run, data about the 3D hand movements of the patient was collected and analysed. The VR activity station features a library of engaging activities and “games” that are simultaneously entertaining for the patient and beneficial for rehabilitation (figure 2).

Telemedicine based on Skype™ with a camera (software version 2.5, freely available from the internet) was used as a communication tool between the therapist and the personal at the non-hospital environment, offering clinical and technical support.



Figure 1. VR activity station (sensegraphics.se).



Figure 2. Game library

1.3 Subject Assessment

A UE test developed in a previous study [10] was used. The subjects had to move the haptic stylus to different targets in the virtual world (Figure 3). The targets appeared one after the other on the screen and disappeared when pointed at. The target placements (32) in the three dimensional space, were apparently random to the patients, but actually set according to a pre-set kinematic scheme for evaluating purposes. The subject had to move as accurately and quickly as possible to each target. Hand position data (haptic stylus end-point) during each trial was gathered.

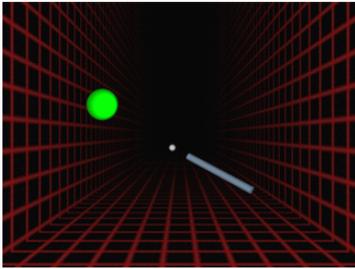


Figure 3. VR-task (screenshot).

The x-, y- and z-coordinates, which were time stamped, gave the basic pattern of hand movement. Time and distance to complete the whole exercise was recorded. From this average velocity and HPR (hand-path ratio - the quotient between actual hand trajectory and the straight-line distance between two targets) was calculated. The basic pattern of stylus movement in space was visualized in Matlab (www.mathworks.com), giving an indication of how hand trajectory and movement quality changed over time. A neglect test developed by our group was used (Figure 4). The VR environment consisted of 20 targets and 60 distracters (2.7 cm diameter). The target was the digit '1' and the distracters were other numbers. The subjects had to press all targets marked with the digit 1, whereby the target changed color, and finally press the red button, indicating when they had finished their search.

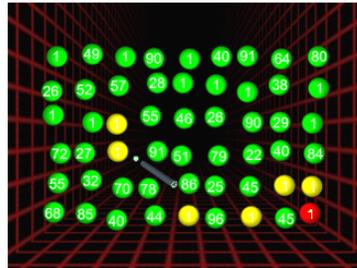


Figure 4. VR-task (screenshot).

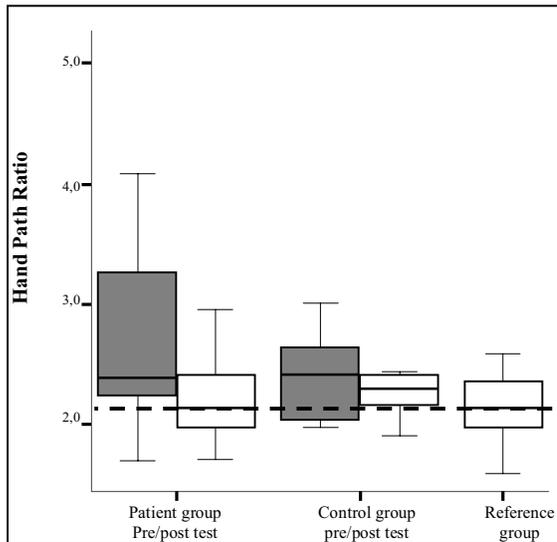


Figure 5. Hand Path Ratio values shown as a boxplot (25th and 75th percentiles) for the treatment (n=16), control (n=11) and reference group (n=11). The median value (thick black line) with 10th, 90th percentiles are shown at the end of the lines.

----- = Reference median values

2. Results

A comparison between pre/post testing suggests that all patients had a higher median in average velocity (m/s) and a decrease in median for the time (s) and HPR parameters as compared with baseline. Figure 5 gives the box plot for HPR parameter.

A further aim was to use the VR activity station as a tool for assessing visuospatial neglect, in order to analyse the manual search performance in a virtual cancellation task. Subjects with neglect as well as subjects recovered from neglect showed aberrant search performance in the VR cancellation task, i.e. mixed search patterns (Figure 6, an example), repeated target pressures, ipsilesional start of search and deviating hand movements. The data indicate that this VR task is more informative for deviating search performance.

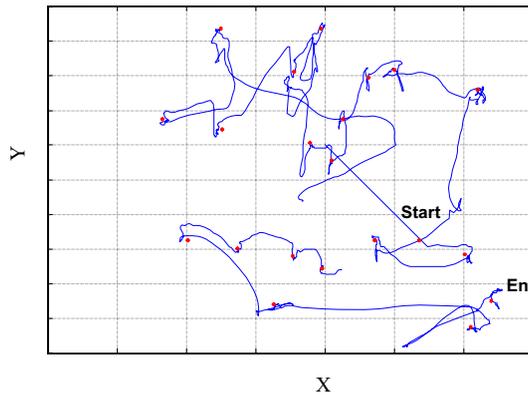


Figure 6. Patient with clinical neglect displays a search mixed pattern.

3. Discussion

Stroke is one of the most common disabling conditions and the need for effective therapies and innovative rehabilitation is clear. The intervention studies show that the subjects made improvements in the kinematic variables measured with the VR system. Further, the VR system was placed outside the hospital to improve access to VR technology by a wider group of stroke subjects. The VR system worked without problems and made it possible to expose persons that otherwise would not have had the chance to try the VR system. Adherence to the programme was excellent and may have been facilitated by the novel technique of VR that would enhance their capacity for better UE function. A reason might be that the persons wanted to try the novel technique of VR with a hope for improving UE function [11]. Attitudes to new technology are affected by the perceived benefits of using it, positive past experiences, quality of information about it, training and follow-up, hands-on experience, the extent to which it meets user needs, and users' enjoyment. Irizarry and colleagues [12] showed that older people welcomed high tech products, including computer-controlled 'smart houses' of the future, but want clearer instructions about how to use them and controls that are easy to read and handle. The use of telemedicine based on Skype™ with a camera was employed as an adjunct to communicate response to the intervention from the clinic to the non-hospital environment. This increased the efficiency while

maintaining a high quality of service to the personal and stroke subjects. When problem arose, i.e. the games would not start, the personal contacted the occupational therapist at the clinic and the problem was solved by remote instructions. Therapy-based rehabilitation is a promising approach in remote training [13, 14]. The computerized neglect test provided a quantitative analysis of detecting small variations in manual search performance otherwise not detected in standard paper-and-pencil tests. It was found that the presence or absence of visual inattention was identified for the same subjects either by using VR or the conventional neglect tests. This is consistent with earlier studies where, in visual search tasks, subjects with neglect not only exhibit omissions of visual targets but also demonstrate more general deficits in their search performance, such as an unsystematic search pattern [15, 16]. Further, both subjects with neglect and subjects who had recovered from neglect showed aberrant search performance in the VR task. The data indicate that this VR task is more informative for deviating search performance as compared with examination of search by more conventional paper-and-pencil tests.

4. Conclusion

The general experience of the VR application approach suggests that this intervention seems to be a promising tool in motor and cognitive rehabilitation, with a wide range of applicability. This project demonstrates that this technology can provide a real-time quantitative 3D task analysis and provides preliminary evidence that interactive computer use with the right training conditions may increase stroke subjects' motor and cognitive skills.

5. Acknowledgements

This research was supported by grants from the Swedish Stroke Victims Association, the Hjalmar Svensson Research Foundation, the Göteborg Foundation for Neurological Research, the Amlöv Foundation, Wennerström's Foundation, the Per Olof Ahl Foundation, SFMI – the Swedish Federation for Medical Informatics, the Federation of Swedish County Councils (VG-region), Trygg-Hansa insurance company, the Swedish Research Council (VR K2002-27-VX-14318-01A and VR 2005-27-IT-15345) and VINNOVA (2004-02260).

References

- [1] National Health Care Quality Register in Sweden. *Analysrande Rapport Riks - Stroke 2004*. [cited 26 February 2008; available from: <http://www.riks-stroke.org>].
- [2] G. Greveson, O. James, Improving long-term outcome after stroke--the views of patients and carers, *Health Trends* **23** (1991) (4), 161-162.
- [3] A. Drummond, Leisure activity after stroke, *Int Disabil Stud* **12** (1990) (4), 157-160.
- [4] U. Sveen et al., Association between impairments, self-care ability and social activities 1 year after stroke, *Disabil Rehabil* **21** (1999) (8), 372-377.
- [5] A.A. Rizzo, G.J. Kim, A SWOT Analysis of the Field of Virtual Rehabilitation and Therapy, Presence: Teleoperators and Virtual Environments **14** (2005) (2) 119-146.
- [6] K. Baheux et al., Diagnosis and rehabilitation of hemispatial neglect patients with virtual reality technology, *Technol Health Care* **13** (2005) (4), 245-260.

- [7] A. Bardorfer et al., *Upper limb motion analysis using haptic interface*. IEEE ASME Trans Mechatron **6** (2001), 253-260.
- [8] R. Boian et al., Virtual reality-based post-stroke hand rehabilitation, *Stud Health Technol Inform* **85** (2002), 64-70.
- [9] R. Kizony et al., Video-capture virtual reality system for patients with paraplegic spinal cord injury, *J Rehabil Res Dev* **42** (2005), 595-608.
- [10] J. Broeren, M. Rydmark, K.S. Sunnerhagen, Virtual reality and haptics as a training device for movement rehabilitation after stroke: a single-case study, *Arch Phys Med Rehabil* **85** (2004), 1247-1250.
- [11] R.N.Barker, S.G. Brauer, Upper limb recovery after stroke: the stroke survivors' perspective, *Disabil Rehabil* **27** (2005), 1213-1223.
- [12] C.A. Irizarry, A. Downing, C. Elford, Seniors-on-line: introducing older people to technology, *Australas Phys Eng Sci Med* **20** (1997), 39-43.
- [13] L. Legg, P. Langhorne, Rehabilitation therapy services for stroke patients living at home: systematic review of randomised trials, *Lancet* **363** (2004), 352-356.
- [14] J.M. Winters, J.M. Winters, A telehomecare model for optimizing rehabilitation outcomes, *Telemed J E Health* **10** (2004), 200-212.
- [15] M. Behrmann, P. Ebert, S.E. Black, Hemispatial neglect and visual search: a large scale analysis, *Cortex* **40** (2004), 247-263.
- [16] A. Parton, et al., Space re-exploration in hemispatial neglect, *Neuroreport*, **17** (2006). (8), 833-836.

An Easy to Use and Affordable Home-Based Personal eHealth System for Chronic Disease Management Based on Free Open Source Software

Tatjana M. BURKOW^{a,1}, Lars K. VOGNILD^b, Trine KROGSTAD^a, Njål BORCH^b, Geir ØSTENGEN^a, Astrid BRATVOLD^c and Marijke Jongsma RISBERG^c

^a *Norwegian Centre for Telemedicine, UNN, Tromsø, Norway*

^b *Northern Research Institute - Norut, Tromsø, Norway*

^c *University Hospital of North-Norway, Tromsø, Norway*

Abstract: This paper describes an easy to use home-based eHealth system for chronic disease management. We present the design and implementation of a prototype for home based education, exercises, treatment and following-up, with the TV and a remote control as user interface. We also briefly describe field trials of the system for patients with COPD and diabetes, and their experience with the technology.

Keywords: Telehealth, Systems Architecture, Patient Monitoring, Disease mg., Telemedicine, User-computer interface, Open source, Computer-supported training, Online/distance education, eHealth.

Introduction

The number of people with chronic diseases such as COPD, diabetes, heart disease, stroke, obesity and cancer increases rapidly. Accounting for 59% of the 57 million deaths annually and 46% of the global burden of disease, chronic diseases are the major cause of death and disability worldwide. It is even projected a 17% increase in deaths due to chronic diseases from 2005 to 2015 [1,2]. In addition, as we live longer and the risk of chronic conditions increases with age, the number of people with one or more chronic condition also increases [3,4,5]. This leads to huge burdens on the healthcare system and require innovative approaches to provide out-of-hospital treatment. The home environment can be a potential arena for treatment and following-up; reducing in-hospitals stay's and costs. Home based eHealth systems can give life and health improvements such as stabilization of the chronic disease, prevention of exacerbations,

¹ Corresponding Author: Tatjana M. Burkow, Norwegian Centre for Telemedicine, University Hospital of North-Norway, P.O. Box 35, N-9038 Tromsø, Norway; E-mail: tatjana.m.burkow@telemed.no

less hospitalization, and lost work days, reduced risk for hospital infections, better treatment compliance, cognitive status, mastering of the disease and patient satisfaction, and improved quality of life [6,7].

A challenge for home based eHealth is that large user groups experience technological barriers today, especially in the elderly part of the population [8,9]. Since the number of chronic conditions *increases* with age, and the computer skills in the population *decreases* with age, there is a need for easy to use technologies. However, many advanced telemedicine systems for chronic patients use computers or smartphones, which are not necessarily a familiar interface for the elderly.

Our home-based eHealth system for chronic disease management is flexible, comprehensive, affordable, easy to use and build on free open source software. We use the TV and remote control as interface in order to achieve high usability.

Having a patient centric view, our system aims at assisting the individuals in the management of their chronic disease, and including them in the healthcare process. It is not a traditional patient remote monitoring system [10,11] where the patient is under surveillance and information typically transferred to a central server to be viewed by healthcare personnel only. The rest of the paper describes our home based system for chronic disease management.

1. Materials and method

1.1. System Design

The patients use their ordinary TV and a remote control as an interface to the system. In addition they need a camera and a headset for audiovisual communication with healthcare personnel and peers suffering from the same disease. The healthcare personnel use their PC-based systems and stand alone videoconferencing systems.

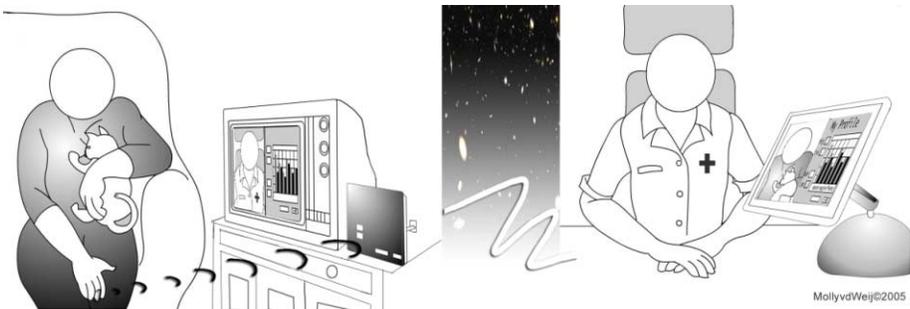


Figure 1: The Residential Patient Device connected to a TV

As part of our system, a *Residential Patient Device* (a dedicated computer) is connected to the users TV and a broadband connection. It provides a rich set of functions to implement personal eHealth-services, provides storage, and a secure and reliable communication channel between the home and the public health service.

Information is visualised on the TV in a useful format for the chronically ill, while the information needed by the healthcare personnel will be visualised on their PC-

based systems in an appropriate format. The following functionality is currently provided by the system:

- Presentation of health related educational material, both video and text-based
- Two-part and multipart videoconferencing
- A *Patient Health Diary* with:
 - Disease related questions
 - Vital signs measurements such as heart rate, oxygen saturation, and blood glucose values and other sensor data (wireless transmitted or manually registered)
- Visualization of information in the Patient Health Diary with possible different views for the patients and the care persons.
- User authentication and data encryption/decryption
- Message transmission

This functionality can be used to orchestrate home-based services for the chronically ill, tailored to the user's needs and abilities. The actual functionality needed will depend on the disease in question and the eHealth service delivered to the users in their home environment.

Adhering to our patient centric view we keep the Patient Health Diary stored in the Residential Patient Device. In this way the information is easily available to the patients at home and not "lost" into the information space of healthcare e.g. the EHR, which traditionally makes it difficult, if not impossible, for the patients to get access to their information later on. This approach also allows the chronically ill to register more information in a self management perspective than otherwise would be possible in a hospital maintained EHR. However, both locally and remote stored educational material can be accessed and visualised by the Residential Patient Device.

Secure message transmission is used to transfer information in the Patient Health Diary to the Hospital Information System (HIS). The patients explicit select and authorize the submission of information to the HIS – keeping the patient in control. However, the Residential Patient Device can be configured to also support remote access by healthcare personnel, giving us flexibility when configuring the system.

1.2. Security and Privacy

Information is stored encrypted in a database on the Residential Patient Device, and we use encrypted message transmission, and audio and video conferencing over an encrypted virtual private network (VPN) connection.

Authentication at home is based on pin codes, and if the system has not been used for a predefined period of time, the user is logged out, and has to enter the pin code again. This is done in order to avoid non-authorized persons to get access. Password authentication is used at the hospital for the healthcare workers.

1.3. Implementation

The current system is implemented on a form factor PC platform running Linux and following a basic Web architecture, where most of the application functions and logic are provided from a local web application server. The user interacts with the application through a standard web browser and separate applications for video

viewing and video conferencing. Patient generated information is encrypted and stored in a local SQL database, from which data may be aggregated and transmitted using the hospital messaging platform.

For the setup with TV and remote control we configure the computer to start up with the web browser in full screen mode, render the information especially for the TV, and use a browser plug-in for mouse-less browsing that enables automatic link numbering (recommended by the BBC style guide for interactive TV). In addition we provide the user with easy switching between browsing, video viewing and videoconferencing.

1.4. Services for COPD and Diabetes and the Trials

To validate our approach and prototype we designed home based services for education, exercises and following-up for both COPD and diabetes.



Figure 2: A field trial participant during individual consultation at home

The COPD services are based on the COPD rehabilitation program at the University Hospital of North-Norway. The patients participated in weekly group-based TV-meetings with education and discussions. One or more thematic educational videos were to be watched before these meetings. Every week they also participated in a group based exercise sessions on TV. They registered daily information in the Patient Health Diary (questions, hart rate and oxygen saturation), and this was used as a basis for a weekly individual TV-consultation with healthcare personnel. In addition the Patient Health Diary was used for self management.

The diabetes services are based on the patient education and consultation services for Type 2 diabetes at the University Hospital of North-Norway. As with COPD, the patients participated in weekly TV-meetings with healthcare personnel and peers, and educational videos were to be seen before the TV-meeting. Their Patient Health Diary (questions and blood glucose values) was used as a basis for the individual TV-consultations, in addition to be used for self management.

Both the COPD and diabetes participants submitted last week's entries in their Patient Health Diary to the healthcare personnel by explicit action each week prior to their consultation.

In order to evaluate the prototype and services we ran a field trial lasting two months where five users with COPD and five with diabetes participated from their own homes. In addition nurse, physiotherapist, doctor and dietician participated from the healthcare site. A qualitative method was used in the evaluation, and patients and healthcare personnel were interviewed after the trials.

2. Results and Discussion

Usability was our main motivation when choosing the TV and remote control as user interface in our current prototype. All of the participating trial patients found the TV and remote control very easy to use, both those who have been using a PC before and the non-PC users. Several of the informants expressed: “*very easy to use*”, “*easy to navigate*”, and “*no problems*”. They also found it easy to both enter and retrieve data in the Patient Health Diary, and the use of educational information and videos. Other studies on TV based telehomecare also reports on high user acceptance [12,13].

The need for providing advanced services through an easy to use TV interface, instead of using a general personal computer, is supported by a study on computer usage and skills in Europe performed in 2005. This shows that 65% in age group 55 – 74 have “no basic computer skills”, and 14% have “low level computer skills”. Even in Norway the corresponding numbers are 30% and 30% respectively [9].

We experienced that it is possible to build an easy to use e-health system out of mostly free open source software (FOSS). Using a PC platform and web technology we implemented several functions for the TV that achieved high user acceptance.

Another approach is the traditional digital interactive TV-based systems. However, they often use technology that is quite expensive, builds on special hardware, and requires proprietary software [14]. Such systems could be difficult to customize for a specific service, limited possibility to expand with new hardware components, and unable to integrate with open source systems. Our technological platform is PC based, where it is much easier to add hardware components, and where free open source software can be used. Using FOSS gave us access to existing software that we could customise and configure to our specific need and setup. Also we can easily use new open source components, for example for EHR records, such as OpenEHR.

Another reason is of course the price (0,-). Healthcare systems are well known to be expensive but free open source software offers an opportunity to challenge this and to make affordable systems.

3. Conclusions

We have shown that it is possible to make a flexible and low cost home based eHealth system using affordable components and mostly free software. By using Internet technologies the Residential Patient Device can be accessed from mobile phones, surfpads or laptops – in addition to the TV.

The evaluation shows high user acceptance from the patients who found the TV and remote control very easy to use; - it was easy to navigate, seek information, easy to enter data into the diary and look it up, and also easy to retrieve and view the educational material.

The current technology is under market validation in Better Breathing, an eTen project. Future work includes the support for wireless and mobile solutions, advanced processing of sensor information and decision support. We will also focus on prevention and disease management for people with more than one chronic condition (comorbidity). This work is performed in MyHealthService, a project in the Tromsø Telemedicine Laboratory, a Centre for Research-based Innovation funded by the Research Council of Norway.

4. Acknowledgements

We would like to thank the rest of the project participants from the Norwegian Centre for Telemedicine, University Hospital of North-Norway, Norut and Well Diagnostics for their contribution, as well as all the patients who participated in the trial. We would also like to thank Luis Fernández Luque, for careful reading and valuable comments to this paper. This work received a grant from the Research Council of Norway (HØYKOM).

5. References

- [1] World Health Organization, Global Strategy on Diet, Physical Activity and Health, Facts related to chronic diseases , 2007, [Online]. Available: <http://www.who.int/dietphysicalactivity/publications/facts/chronic/en/index.html>, [Accessed 13 November 2007].
- [2] World Health Organization, Preventing chronic diseases: a vital investment WHO global report, 2005, [Online]. http://www.who.int/chp/chronic_disease_report/content/en/index.html, [Accessed 13 November 2007].
- [3] World Health Organization, Chronic Diseases and Health Promotion, Chronic diseases and their common risk factors, Information sheet, 2007, [Online]. Available: http://www.who.int/chp/chronic_disease_report/media/Factsheet1.pdf, [Accessed 13 November 2007].
- [4] Centre for Disease Control and Prevention and The Merck Company Foundation. The State of Aging and Health in America 2007. Whitehouse Station, NJ: The Merck Company Foundation; 2007. [Online]. Available: <http://www.cdc.gov/aging/> [Accessed 13 November 2007].
- [5] Mannina, D.M., Buist, A.S., Global burden of COPD: risk factors, prevalence and future trends, *The Lancet*, Vol 370, Sept. 1 2007.
- [6] Noel, H.C., Vogel, D.C., Erdos, J.J., Comwall, D., Levin, F. Home Telehealth Reduces Healthcare Costs. *Telemed Journal and e-Health*. 2004, 10(2): 170-183.
- [7] Casas, A., Troosters, T., Garcia-Aymerich, J., Roca, J., Hernandez, C., Alonso, A., Pozo, F. del, Toledo, P. de, Anto, J. M., Rodriguez-Roisin, R. et al., Integrated care prevents hospitalisations for exacerbations in COPD patients, *Eur. Respir. J.*, July 1, 2006; 28(1): 123 - 130.
- [8] Eurostat, The digital divide in Europe - Issue number 38/2005. [Online]. Available: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-NP-05-038/EN/KS-NP-05-038-EN.PDF [2007, November 13].
- [9] Eurostat, How skilled are Europeans in using computers and the Internet? - Issue number 17/2006, [Online]. Available: flexible adaptable education system with changing information sources http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-NP-06-017/EN/KS-NP-06-017-EN.PDF , [Accessed 13 November 2007].
- [10] deToledo, P.; Jimeanez, S.; delPozo, F.; Roca, J.; Alonso, A.; Hemandez, C., Telemedicine Experience for Chronic Care in COPD, *IEEE Transactions on Information Technology in Biomedicine*, Volume 10, Issue 3, July 2006 Page(s): 567 – 573.
- [11] Starren, J., Hripcsak, G., Sengupta, S., Abbruscato, C.R., Knudson, P.E., Weinstock, R.S., Shea, S., Columbia University's Informatics for Diabetes Education and Telemedicine (IDEATel) Project Technical Implementation, *J Am Med Inform Assoc*. 2002 Jan–Feb; 9(1): 25–36.
- [12] Guillen, S., Arredondo, M.T., Traver, V., Garcia, J.M., Femandez, C., Multimedia telehomecare system using standard TV set, *IEEE Transactions on Biomedical Engineering*, Volume 49, Issue 12, Dec 2002 Page(s): 1431 – 1437.
- [13] Karagiannis, G.E., Stamatopoulos, V.G., Kotis, T., Roussos, G., Gatzoulis, M.A., Health and lifestyle management via interactive TV in patients with severe chronic cardiovascular diseases, *Journal of Telemedicine and Telecare*. Volume 12, Supplement 1, July 2006, pp. 17-19(3).
- [14] Maglaveras, N., Chouvarda, I. et al., Citizen Centered Health and Lifestyle Management via Interactive TV: The PANACEIA-ITV Health System. *AMIA Annu Symp Proc*. 2003; 2003: 415–419.

Aligning Lay and Specialized Passages in Comparable Medical Corpora

Louise DELEGER^{a1} and Pierre ZWEIGENBAUM^b

^a*INSERM UMRS 872 Eq. 20, Paris, France*

^b*LIMSI-CNRS, F-91403 Orsay, France*

Abstract. While the public has increasingly access to medical information, specialized medical language is often difficult for non-experts to understand and there is a need to bridge the gap between specialized language and lay language. As a first step towards this end, we describe here a method to build a comparable corpus of expert and non-expert medical French documents and to identify similar text segments of lay and specialized language. Among the top 400 pairs of text segments retrieved with this method, 59% were actually similar and 37% were deemed exploitable for further processing. This is encouraging evidence for the target task of finding equivalent expressions between these two varieties of language.

Keywords. Natural Language Processing, Consumer Vocabulary, Comparable Corpora

Introduction

Over the years the general public has been dealing with an increasing amount of medical information, whether from the Internet or from gaining access to their own medical records as legislations allow them to. However this information is not always expressed in a language suitable to lay people. Specialized medical language is indeed difficult for non-experts and patients may not understand the information that is presented to them [1,2]. There is therefore a real need to empower patients with means which facilitate their understanding of medical language [3,4].

Previous work has addressed the issue in several ways. [5] developed methods to help lay people query the Web for health information. A number of papers emphasized the need for consumer health vocabularies [6,7] and investigated methods to this end [8]. [9] identified and defined difficult medical terms. Building a lexicon of linked specialized medical terms and lay expressions was also investigated as a way to bridge the gap between the two types of languages [10]. A preliminary step to many applications is to examine characteristics of medical texts [11,12]. [13] used such characteristics to automatically categorize expert vs. non-expert Web pages in Russian and French.

¹ Corresponding Author: Louise Deléger, INSERM UMRS 872 Eq. 20, 15 rue de l'École de Médecine, 75006 Paris, France; E-mail: louise.deleger@spim.jussieu.fr

We aim at characterizing patient-friendly documents as opposed to expert documents in order to help design ways to adapt specialized medical information to patients. To do so we propose to build a comparable corpus of lay and specialized medical documents and to study the relations between the two sides of this corpus. More specifically, we would like to find correspondences between the two varieties of languages—lay and specialized—in these documents, in the same line as [10] but for French. This raises two issues: how to collect such a corpus (Section 1.1) and how to process it to identify corresponding passages (Section 1.2) in a setting where, as we shall see below, the comparability of documents is less ideal than in [10]. We then expose the evaluation conducted (Section 1.3) and its results (Section 2), discuss them and conclude (Section 3).

1. Material and Methods

1.1. Acquisition of a Comparable Corpus of Specialized and Lay Texts

We chose to work with medical documents dealing with the topic of tobacco and nicotine addiction and built a corpus of such documents in French, containing texts intended for experts and texts intended for the general public. *Comparable* corpora usually refer to multilingual text collections that address the same topic without being translations of each other. Here the notion is applied to monolingual texts from different genres—lay and specialized—but dealing with the same topic.

Building a corpus is far from being a trivial or an immediate task. A corpus must suit the needs and objectives of the target task, in terms of both size and quality. Today, a popular way of acquiring a corpus is collecting it from the Web [14], as it provides easy access to a virtually unlimited number of documents. When dealing with a Web corpus several issues arise. The first issue is the relevance of the retrieved documents to the targeted domain and is highly dependent on the method used to gather the documents. Several methods can be used for collecting a corpus from the Web on a specific topic: (1) general-purpose search engines, such as Google, can be queried with selected key words [15]; (2) domain-specific search engines can be used (in domains where they exist); (3) direct retrieval of relevant Web pages is also possible when these are known to the user (either directly or through a quick glance at the output of a Web search).

Another important issue specific to our type of corpus is the relevance to the targeted genre. We are gathering a comparable corpus of two genres—lay vs. specialized. Hence the need to classify each collected document as belonging to one genre or the other. Two main approaches are possible: either (a) relying on automatic text categorisation methods [16] (this type of approach is automatic but may introduce potential noise caused by categorization errors); or (b) using resources that include a classification of the documents or resources known to belong to one particular category.

We used methods belonging to points (2), (3) and (b). We discarded generic search engines since the documents would not be categorised and relevance to the domain would be especially questionable with a wide-range domain such as tobacco. We decided in favour of a more restricted search involving a small part of manual work.

First we queried two health search engines that we knew of (the health web portals CISMef² and HON³) with a list of key words. Both provide access to trustworthy Web pages and allow the user to search for documents targeted to a population (e.g., patient-oriented documents). We also knew of relevant websites. Those were French governmental websites, including that of the HAS⁴ which issues guidelines for health professionals, and that of the INPES⁵ which provides educational material for the general public; as well as health websites dedicated to the general public, including Doctissimo⁶. The previous search results also allowed us to identify two more websites dealing with nicotine addiction and aiming at the general public: Tabac Info Service⁷ and Stop-tabac⁸.

Once collected, a corpus needs to be cleaned and converted into an appropriate format for further processing—i.e., extracting the textual content. HTML documents (as we have in our corpus) contain a certain amount of irrelevant information such as navigation bars, footers and advertisements—referred to as “boilerplate”—which can generate noise. While getting rid of HTML code is easy, identifying boilerplate is much more challenging [17]. Boilerplate removal methods can rely on HTML structure, on visual features (placement and size of blocks) and on plain text. We made use of the HTML structure in several ways. First, we noticed that some HTML documents contained meta-information as to the location of the content text in the pages—for instance, a block may have the attribute “*class=content*”, or comments such as “*<!-- content start-->*” may be present. We then used HTML structure to identify potential navigation bars: we made the assumption that any list of links at the beginning of documents with a small amount of text had a high probability of being a navigation bar, so we removed such lists. Finally, we relied on the density of HTML tags to identify the start and end of content text in a document: we measured the ratio of text throughout the document in a sliding window of ten words. We defined the starting point for content extraction as the first location where the ratio became equal to 80% and the end as the last location where the ratio became equal to 80%. We also relied on plain text features to try to spot obviously irrelevant text, such as phone and fax numbers and e-mails (as often appear at the end of documents).

1.2. Identification of Relations between Specialized and Lay Texts

Our long-term objective is to find equivalent expressions in specialized and lay texts. As a first step, we tried to relate text passages taken from both sides of our comparable corpus which address similar topics and might thus contain corresponding expressions.

1.2.1. Topic Segmentation

Texts rarely constitute homogeneous units, as multiple topics are usually addressed in a single text. It might therefore be difficult to relate two entire texts, so we chose to work at the sub-level of topic segments. We performed topic segmentation on each text using the segmentation tool TextTiling [18]. A segment may correspond to one or several

² <http://www.cismef.org/>

³ <http://www.hon.ch/>

⁴ <http://www.has-sante.fr>

⁵ <http://www.inpes.sante.fr/>

⁶ <http://www.doctissimo.fr/>

⁷ <http://www.tabac-info-service.fr/>

⁸ <http://www.stop-tabac.ch/>

paragraphs. Another possibility would have been to work at the paragraph level or even at the sentence level—this provides alternative ways to explore in further work.

1.2.2. *Assessing the Similarity of Text Segments*

Once our corpus was topically segmented, we tried to identify pairs of text segments addressing similar topics. For this we used common, vector-based measures of text similarity: Cosine and Jaccard. Both measures give a similarity score ranging from 0 to 1. We computed them for each pair of topic segments in the cross-product of both corpus sides. The input segments were transformed into vectors of words and their similarity was established based on the proximity of the vectors. Given two vectors A

and B, $Jaccard(A; B) = \frac{\|A \cap B\|}{\|A \cup B\|}$ and $Cosine(A; B) = \frac{A \cdot B}{\|A\| \cdot \|B\|}$. Cosine takes into

account the frequency of the words while Jaccard only relies on the presence or absence of a word.

1.3. *Evaluation*

We evaluated the similarity results by selecting a sample (one every four) from the 600 pairs of text segments with the highest similarity scores, for each similarity measure. We obtained two samples of 150 pairs to review, one corresponding to the Cosine measure and one to the Jaccard measure. We evaluated them against the following two criteria: (1) whether the text segments were actually similar; (2) whether the text segments contained equivalent expressions, which a future step will aim at detecting automatically—i.e., whether they were exploitable for further processing.

2. Results

Table 1 shows the size of our corpus. Both sides (lay and specialized) of the corpus contain approximately the same number of words but the number of lay documents is far greater. This means that our search brought more lay documents, but also that expert texts are much longer. Topic segmentation resulted in 3,226 segments in specialized texts and 2,769 in lay texts (see table 2).

Evaluation results for the most similar segment pairs are given in figure 1. We see that the proportion of similar pairs and of exploitable pairs is rather good up to the 400th pair (59% and 37% with Cosine) which slowly decreases afterwards, for both similarity measures. This shows that we may draw the limit of similar pairs to the first 400 ones, and discard the other pairs. The proportion is also almost always higher with Cosine which seems to indicate that this measure is more appropriate. A prospect for assessing the similarity would be to use a combination of those two measures as a similarity score, instead of just one or the other.

Table 1. Size of acquired corpus

	Documents	Words
Specialized	61	580,945
Patient-friendly	546	640,401
Total	607	1.22M

Table 2. Number of segments produced by topic segmentation

	Specialized	Patient-friendly	Total
Nb of segments	3,226	2,769	5,995

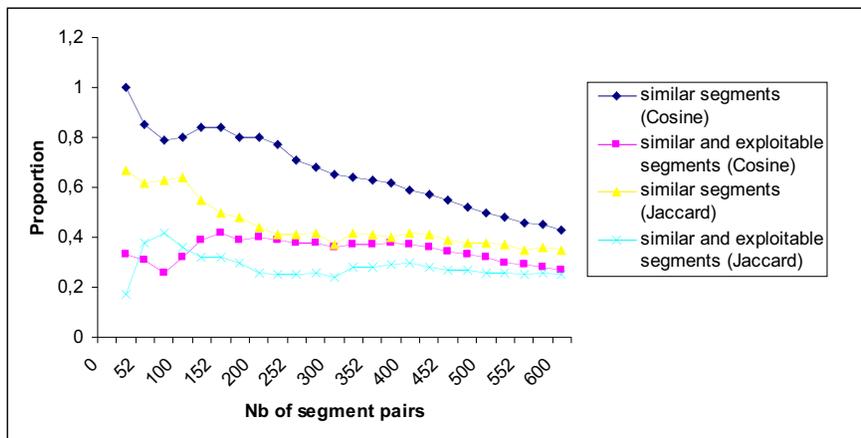


Figure 1. Results for the 600 best segment pairs

Table 3. Excerpt from a pair of text segments

Specialized	Lay
Le tabagisme de l'homme serait à l'origine : 1. d'une dysfonction érectile réversible à l'arrêt ; ...	Les effets délétères du tabac sur la fertilité en général : Chez l'homme : Dysfonction érectile (qui disparaît à l'arrêt du tabac) ; ...

An excerpt of similar text segments is given in table 3 (equivalent expressions are highlighted in corresponding colors). Similar but non-exploitable pairs include navigation bars that could not be filtered out during conversion from HTML to text (hence the need to clean the corpus as much as possible after collecting it), and portions of texts describing statistics.

3. Discussion and Conclusion

We built a comparable corpus of lay and specialized medical documents in French from which we were able to identify pairs of similar text segments that seemed exploitable for finding equivalent expressions between lay and specialized language. This is encouraging evidence since these documents, although addressing the same general topic of tobacco and nicotine addiction, were a priori rather different since they come from various sources and are targeted to different populations.

The number of such pairs of similar segments may seem rather small compared to all potential segment pairs. However, as above mentioned, we are dealing with a comparable corpus of a priori rather dissimilar documents and not with a parallel

corpus, in contrast to [10] whose corpus had paired lay and specialized texts. This also raises the issue of the corpus size needed to provide sufficient coverage of the domain. The limited number of selected text segments may indicate that we need to build a larger corpus.

Future work includes finding equivalent expressions (sentences, phrases, words) among the pairs of text segments; testing a different similarity measure that would be a combination of the two measures used in this work; segmenting the texts into paragraphs or into sentences instead of topic segments.

Finally we can wonder whether results would be different with a different topic than tobacco. Testing the method on a different corpus is also a prospect of this work.

To sum up, we described a method to build a corpus of expert and non-expert medical French documents and identify similar text segments of lay and specialized language, as a first step towards bridging the gap between these two varieties of language. We were able to identify similar text segments. Although there is room for improvement and further testing, this gives good hope for our target task of finding equivalent expressions between lay and specialized medical language.

References

- [1] Lerner EB, Jehle DV, Janicke DM, and Moscati RM. Medical communication: Do our patients understand? *Am J Emerg Med* Nov 2000;18(7):764–6.
- [2] Chapman K, Abraham C, Jenkins V, and Fallowfield L. Lay understanding of terms used in cancer consultations. *Psychooncology* 2003;12(6):557–66.
- [3] McCray AT. Promoting health literacy. *J Am Med Assoc* Mar-Apr 2005;12(2):152–63.
- [4] Borin L, Grabar N, Gronostaj MT, et al. Empowering the patient with language technology. Technical report Work Package 27 Deliverable 2, Semantic Mining Network of Excellence, 2006.
- [5] Zeng QT, Crowell J, Plovnick RM, et al. Assisting consumer health information retrieval with query recommendations. *J Am Med Assoc* Jan-Feb 2006;13(1):80–90.
- [6] Zeng QT and Tse T. Exploring and developing consumer health vocabularies. *J Am Med Assoc* 2006;13(1):24–9.
- [7] Zielstorff RD. Controlled vocabularies for consumer health. *Biomed Inform* 2003;36(4-5):326–3.
- [8] Zeng Q, Tse T, Divita G, et al. Term identification methods for consumer health vocabulary development. *J Med Internet Res* 2007;9(1):e4
- [9] Elhadad N. Comprehending technical texts: Predicting and defining unfamiliar terms. In: *AMIA Annual Symposium*, Washington, DC. 2006:239–43.
- [10] Elhadad N and Sutaria K. Mining a lexicon of technical terms and lay equivalents. In: *Proceedings of ACL BioNLP Workshop*, Prague, Czech Republic. 2007:49–56.
- [11] Leroy G, Eryilmaz E, and Laroya B. Health information text characteristics. *Proc AMIA Annu Symp* 2006:479–83.
- [12] Zeng-Treitler Q, Kim H, Goryachev S, et al. Text characteristics of clinical reports and their implications for the readability of personal health records. *Medinfo* 2007;12:1117–21.
- [13] Grabar N, Krivine S, and Jaulent MC. Classification of health webpages as expert and non expert with a reduced set of cross-language features. *Proc AMIA Annu Symp* 2007: 284–88.
- [14] Kilgarriff A and Grefenstette G. Introduction to the special issue on the web as corpus. *Computational Linguistics* 2003;29(3):333–47.
- [15] Ghani R, Jones R, and Mladenic D. Using the Web to create minority language corpora. In: *Tenth International Conference on Information and Knowledge Management*, 2001.
- [16] Sebastiani F. Machine learning in automated text categorization. *ACM Computing Surveys* 2002;34(1):1–47.
- [17] Baroni M and Sharoff S. Creating specialized and general corpora using automated search engine queries. In: *Web as Corpus Workshop*. Birmingham University, UK, Jul 2005.
- [18] Hearst M. Texttiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics* 1997;23(1):33–64.

Knowledge Engineering as a support for building an Actor Profile Ontology for Integrating Home-Care Systems

Karina GIBERT^a, Aida VALLS^b, David RIAÑO^c

^a *KEMLG research group, Universitat Politecnica de Catalunya
Dept. of Statistics and Operational Research, Barcelona, Spain*

^b *ITAKA research group, ^c Research group on AI,
Universitat Rovira i Virgili*

Dept. of Computer Science and Mathematics, Tarragona, Spain

Abstract. One of the tasks towards the definition of a knowledge model for home care is the definition of the different roles of the users involved in the system. The roles determine the actions and services that can or must be performed by each type of user. In this paper the experience of building an ontology to represent the home-care users and their associated information is presented, in a proposal for a standard model of a Home-Care support system to the European Community.

Keywords. Home-based eHealth, Decision Support and Knowledge Management, Health Information Systems, Telehealth, Distributed Systems, Knowledge-based systems, Concept representation-preservation, Home-care.

Introduction

K4CARE is a project financed by the European Commission devoted to develop an intelligent web platform to provide e-services to health professionals, patients and citizens involved with the Home Care (HC) of elderly patients living at home. Those services aim to improve the capabilities of the new EU society to manage and respond to the needs of the increasing number of senior population requiring a personalized HC assistance. From a medical point of view, one of the main goals of the project was to identify which were the common and basic home care structures shared by the main sanitary systems in Europe. They are called HCNS (Home Care Nuclear Services) and they comprise the minimum elements needed to provide a basic HC assistance. From a technological point of view, the K4CARE project aims to provide an intelligent platform which can support HCNS management and also, possible later extensions to specialized services such as those coming from Oncology or Rehabilitation units ([1],[2]), these are called HC Accessory Services (HCAS).

To build a platform that supports such a complex HC model is not a trivial task and requires a careful design. In the K4CARE project, the platform is implemented using a multiagent system, which facilitates the interaction between different types of

wireless devices, the communication between different components and the distribution of tasks according to the medical roles. Another critical part of the construction of such a complex system is the design of the information system storing the knowledge which describes the behaviour of the platform. In this project this has been modelled using ontologies [3], a formalism widely used in Knowledge Engineering, AI and Computer Science, which provide a formal frame to represent the knowledge related with a complex domain, as a qualitative model of the system.

However, to build the ontology of a phenomenon as the one faced by the K4CARE project is a huge and a very complex task, because a great number of services, people and institutions are involved in the HC, and it is very difficult to find a formal representation to express all the interactions among the components of the system. The medical knowledge that the K4CARE platform has to manage is represented in different Knowledge Sources, due to their different functionalities in the system [4]. One of those knowledge sources is called APO (Actor Profile Ontology) and it is devoted to contain all the knowledge about the different kind of actors involved in a HC assistance system and their potential functionalities and possible interactions. Thus, APO is defining the knowledge behind the global behaviour of the system and the services provided. The purpose of the APO is to facilitate the integration and coordination of the different actors that are needed to assure a good HC assistance, since the APO clearly defines what and how care will be done in this new HC model.

Describing the design, the contents, and the role of the APO are the focus of this paper, as well as to insist on the problematic related with the construction of this kind of ontologies when real systems that support very complex real domains are constructed. A deeper understanding of the robust, consistent and reliable management of APO comprises a methodology and several tools focused in another paper ([5]).

In the following section, we describe the contents of an ontology and a formal methodology for building ontologies. Section 2 is devoted to explain how this methodology has been applied to build the APO and the results obtained at each stage, including the main details about the contents of the APO. Finally, section 3 highlights the main aspects of this work and presents the conclusions.

1. Materials and methods

Ontologies

As said before, the APO is an ontology. In [3] ontology is defined as a formal, explicit specification of a shared conceptualization, which provides an abstract model of some phenomenon by identifying the relevant concepts of that phenomenon. Ontology captures the common knowledge, that is, not a personal view of the target phenomenon of some particular individual, but one accepted by a group. In our case, this group is formed by the different medical partners of the K4CARE project, which include people from eastern and western European countries. There are different ontology knowledge representation formalisms, all of them sharing 3 common components:

- **Classes:** they represent concepts taken in a broad sense.
- **Relations:** they represent a type of association between concepts of the domain. Binary relations are sometimes used to express concept attributes.
- **Instances:** they are used to represent elements or individuals in the ontology.

Ontology Engineering

The set of activities that concern the ontology development process, the ontology life cycle, the principles, methods and methodologies for building ontologies, and the tool suites and languages that support them is called Ontological Engineering [6]. With regard to methodologies, several proposals exist, among which On-To-Knowledge [7] is the one used here. It is based on five steps:

- Feasibility study, which is the basis for next steps.
- Kickoff: devoted to establish the ontology requirements specification documents. These documents define the domain and the goal of the ontology, the design guidelines, the available knowledge sources, the potential users, and the use cases. Due to the particular complexity of the K4CARE system, and the aim of establishing European standards for K4CARE, the Kickoff is developed by means of a panel of medical and social experts in the field of geriatrics and home care.
- Refinement: it is addressed to obtain a mature ontology in two steps, taking as a basis the initial specification provided in the kick-off.
 - Knowledge elicitation with domain experts. This is a very critical step, since lots of implicit knowledge is unconsciously used by experts in their daily reasoning. This knowledge is crucial for correct modelling and usually does not appear in the kick-off specification. This missing knowledge produce inconsistencies, redundancies or incompleteness that may cause wrong performances of the Ontology
 - Formalization: Once a consistent, non-redundant, and presumably complete description of the domain is clear, proper classes and relations that correctly formalize it have to be identified.
- Evaluation: to prove the correctness and usefulness of the ontologies:
 - Checking the requirements and competency questions
 - Testing the ontology in the target application environment.
- Maintenance: On-To-Knowledge proposes to carry out ontology maintenance as a part of the system software.

2. Results

In this section the most significant results of the ontology building are presented.

Kickoff: The result of the experts panel process is the ontology requirements specification document, where the Nuclear Home Care Services to be supported by the K4CARE platform are specified, together with the actors involved, including all sort of professionals (patient, relatives, citizens, and involved social organisms), the tasks they can perform to complete medical services, the processes where they take part, the documents they will exchange and the kind of access they will have to the information generated as a result of those activities.

Refinement: Two steps compose this phase:

- Elicitation: The interaction between knowledge engineers and medical experts identified some missing elements in the original specification. As a consequence:
 - A detailed specification on how the different services will be performed was made. A procedure was associated to every service indicating which steps

Actor, for individual participants (patient, stable members such as nurse, physician in charge of the patient, head nurse, social worker, family doctor and additional care givers). Entities store information about the services they can initiate and the actions they can perform. Actors also have some rights over Documents they can read or write.

In the K4CARE Model, a **Service** is defined as a HC activity that involves the work of one or more HC actors in a coordinated way. They are classified into Access Services (management services), Patient Care Services, and Information Services. The property `serviceInitiatedBy` informs about the entities that are able to activate the service (i.e. capabilities and liabilities), while `hasProcedure` can indicate one or more alternative ways of performing that service by means of procedures. Finally, `hasDocument` binds documents that may be required in the achievement of that service.

Actions represent the single steps that should be done to perform a service. The ontology distinguishes many subtypes as social action, case management action, back-office action, nursing action, etc. Actions have an object (`hasObject` property) that receives the action (usually a patient), and a subject (`hasSubject` property) that is the entity able to perform that action. Actions can also work over one or several documents (`usesDocument` property).

A **Procedure** is the representation of the way a Service is provided in terms of the available actions and the rest of services. It stores information about the referred service the procedure represents (`isProcedureOf` property), the steps of the procedure (`hasStep` property) or actions and services involved in the provision of the service and the SDA algorithm that defines how the procedure must be applied, that is, which is the flow of actions (property `hasSDAFile`). Different care units can have different SDAs for the same procedure, for this reason the ontology includes also the concept **SDA**.

Documents model the communication between Entities all along Service performance. They can be used in different Procedures (property `isDocumentOf`), and can be used by different Actors with different rights (properties `isWrittenBy`, `isWrittenSometimesBy`, `isReadBy`, `isReadSometimesBy`).

Every leaf class of the hierarchies of concepts Entity, Service, Procedure and SDA inherits also from the **Care Unit Element** hierarchy. This permits to indicate which Care Unit they refer to. For every new HC accessory service (HCAS) used to extend the capabilities of the K4CARE Model, the APO must have a new subclass of Care-UnitElement with the name of the new HCAS (e.g. rehabilitation, oncology).

3. Discussion, conclusions and future work

This research is involved with modelization of domain knowledge for building an intelligent platform supporting the HC assistance of elderly people. The starting point is the proposal elaborated manually by a panel of experts defining the standards of HC in future EU countries [1]. The proposal integrates the best practices of old and new EU countries in a handbook of good medical assistance to ill, disabled, chronic senior patients in a technological society. Defining this standard required a huge effort from the medical partners of the project, who had to formalize the behaviour of a HomeCare system as precisely as possible, in a structured way. This process made explicit valuable medical knowledge that was not available before and that was the input of this work. Expert's proposal in [1] identifies which are the HCNS. This document constituted a general specification at the medical level that required deeper formalization at the engineering level, since some relevant information was missing.

In Knowledge Engineering it is well-known that formalizing complex domains is a very difficult task, mainly because of existence of implicit knowledge, which is rarely expressed in a first expert's specification. Knowledge engineering helps to refine and formally validate the medical model in terms of correctness from a logical point of view. High interaction between experts and knowledge engineers is required in this step and good collaboration was achieved in this project. Non-trivial redundancies, inconsistencies and lack of information needed to complete the specification were identified. The correctness of the APO is critical for a good performance of the system.

Finally, APO was internally organized to mimic the real structure of HC assistance, and its internal architecture was designed to be flexible enough for supporting later enlargements with new specialized care units.

Currently, a first prototype of the K4CARE platform is being implemented and future tests in real scenario are in progress. APO plays a crucial role in the global architecture of the system, since it is defining the tasks that each user is allowed to do. Moreover APO is defining permissions and interactions between the different people involved in the HC assistance, from the patient itself, to the main professionals as family doctors or social workers, what opens the door to having a real integrated assistance in real time. This work is an application of the knowledge engineering methodologies to a real domain and a meaningful scenario describing HC was successfully modeled.

Acknowledgements

This research has been partially financed by the EU Project K4CARE (IST-2004-026968) and the Spanish HYGIA project (TIN2006-15453-C04). J. Casals provided a first implementation of the APO. The medical and social experts in the field of geriatrics and HC participating in the panel of experts were F. Campana, R. Annicchiarico, S. Ercolani, A. Federici, T. Caseri, E. Balint, L. Spiru, D. Amici, R. Jones and P. Mecocci.

References

- [1] Campana, F. et al. Knowledge-Based HomeCare eServices for an Ageing Europe. Deliverable D01, EU (2007), http://www.k4care.net/fileadmin/k4care/public_website/downloads/K4C_Model_D01.rar.
- [2] Casals, J., Gibert, K., Valls, A., Deliverable DO4.1: Sample APOs, K4CARE document, (March 2007).
- [3] Studer, R., Benjamins, R., Fensel, D.: Knowledge engineering: Principles and methods. IEEE Trans. On Data and Knowledge Eng. 25 (1998) 161–197.
- [4] Batet, M., Gibert, K., Valls, A., *A data abstraction layer as knowledge provider for a medical multi-agent system*, 11th Conference on Artificial Intelligence in Medicine (AIME 2007). D. Riaño et al., Eds. Amsterdam, The Netherlands (2007).
- [5] Casals J., Gibert K., Valls, A. Enlarging a medical actor profile ontology with new care units. Proceedings of the I-Workshop From Knowledge to Global Care, 11th Conference on Artificial Intelligence in Medicine (AIME 2007). D. Riaño, et al. Eds. Amsterdam, The Netherlands (2007).
- [6] Gómez-Pérez, M.F.L.: Ontological Engineering. Springer (2004).
- [7] Staab, S., Schnurr, H., Studer, R., Sure, Y.: Knowledge process and ontologies. 12th Banff Workshop on Knowledge Acquisition, Modelling and Management (1999) 4–4:1–20.
- [8] Fensel, D., van Hermelen, F., Horrocks, I., McGuinness, D.L. and Patel-Schneider, P.F.: OIL: An Ontology Infrastructure for the Semantic Web. IEEE Intelligent Systems (16). 2001. 38-44.
- [9] Noy, NF., Ferguson, RW., Musen, MA.: The knowledge model of Protégé-2000: Combining interoperability. In: Dieng R, Corby O (eds) 12th International Conference in Knowledge Engineering and Knowledge Management (EKAW'00). France. LNAI 1937, Springer-Verlag, (2000) 17-32.
- [10] Riaño, D. The SDA Model v1.0: A Set Theory Approach. DEIM Research report DEIM-RT-07-001, (2007), <http://deim.urv.es/receca/reports/DEIM-RT-07-001.pdf>

Evaluation of the Use of Digital Pens for Pain Assessment in Palliative Home Healthcare

Leili LIND

Department of Biomedical Engineering / Medical Informatics, Linköping University, Sweden, and

Santa Anna IT Research Institute AB, Sweden

Abstract. An information system supporting pain assessment in palliative home healthcare was implemented at the hospital-based home care clinic of University Hospital in Linköping, Sweden. Using digital pens and pain diaries, pain assessments were sent from the patients' home to the professional caregiver. A total of 12 patients participated in the study. Patients, spouses and professional caregivers were interviewed. Qualitative content analyses were performed on the study material. All patients managed to use the pain assessment method, they experienced an improved contact with the caregivers and had a sense of increased security. After an initial cautious outlook the caregivers experienced positive outcomes for themselves and their patients. The medical records showed that the method had had impact on treatment. In conclusion, the home healthcare solution provided an effortless method for pain assessment with a high degree of user acceptance for palliative patients and had positive influences on the care.

Keywords. Telehealth, patient monitoring, Home care services; hospital-based

Introduction

Adequate symptom control is one of the most important components of delivering effective palliative care [1-2], whether care is given in a hospital or at home. As people come closer to death their symptoms change rapidly and their need to communicate with their families, friends and health professionals also changes [3]. For cancer patients in a palliative state a systematic assessment of pain is crucial [2, 4], and the American Cancer Society, ACS, [5] has established policies that support palliative care, with particular focus on pain and symptom management. Stone *et al.* has stated the need for frequent assessment of patients' momentary pain to capture pain variations around-the-clock and allow the caregivers to describe the pain pattern over time [6].

Telecommunication and IT can be used to support the patients in assessing and reporting their pain, and such technologies can also support the professional caregivers to work with pain control in a more structured way. Studies where patients have assessed and reported their pain using paper-based and/or electronic pain diaries have been performed earlier [7-9]. Electronic, handheld pain diaries have been found useful

and to provide a high degree of patient satisfaction in studies with patients suffering from chronic pain [8-10]. For an older patient group, pain assessment can be complicated by sensory and cognitive impairment, and motor loss [11]. Older users of handheld computers, such as PDAs and cell phones, often find the screen brightness insufficient and the screen fonts too small to read clearly [12].

In our study, where palliative home care patients suffering from end-stage cancer assessed their pain and reported intake of extra doses of pain relieving medicine by use of a pain diary, we applied digital pen technology from Anoto™¹. Various digital pen technologies have been tried in studies for clinicians' use [13-15], but to our knowledge there were no earlier reported experiences from patients' use of digital pen technology in palliative home healthcare.

The aim of this study was to explore and describe palliative home care patients' experiences of assessing their pain by using paper pain diaries together with digital pen and mobile Internet technology, and furthermore, to explore and describe the professional caregivers' experiences of the system and of their patients' use of this pain assessment method.

1. Material and Methods

The main requirements for the system and study, together with a description of the system, can be found in Lind & Karlsson [16]. Results from the patient/spouse part of the study can be found in Lind *et al.* [17].

The Anoto Technology is based on digital pen and paper. The digital paper is printed with a dot pattern that is almost invisible to the eye. The digital pen used in our study, a Chatpen™, looks and feels like an ordinary ballpoint pen and is used in the same way. The internal part consists of a camera, an image processing unit and a communication unit. The pen knows its exact position in the whole pattern space by using the dots' displacements in the pattern, and the strokes made by the pen are recorded and can be transferred via wireless Internet technology to a server.

Traditionally, patients receiving care from the HBHC clinic had frequent contact with their professional caregivers, either via a telephone or home visits, during which they conveyed their pain and other symptoms. The reporting of pain could be accomplished in a non-structured way through an informal talk about the pain with the visiting nurse/physician or on the phone. Occasionally pain assessment could be performed using a VAS instrument carried by the nurse/physician.

The Regional Ethical Review Board in Linköping gave permission to carry out the study which was conducted at the hospital-based home care (HBHC) clinic, Linköping University Hospital. In-depth face-to-face interviews were performed with patients, spouses and professional caregivers. The study material further consisted of an ease-of-use questionnaire [16], medical records, and the system log.

- In the patient-spouse part a descriptive and explorative case study design with a cross-case content analysis was performed, including interviews with patients and spouses, an ease-of-use questionnaire, medical records and the system log.

¹ Anoto (<http://www.anoto.com>)

- In the professional caregiver part a qualitative descriptive and explorative design with a content analysis approach was performed, including interviews with professional caregivers, the system log and the participating patients' medical records.

2. Results

Evaluation of the pain assessment system aimed at exploring and describing patients' and professional caregivers' experiences of the system. For an overview of socio-demographic and clinical data from the study patients, see Lind *et al* [17].

Six sub-categories showing patients' experiences of using a pain diary and digital pen technology for frequent pain assessment were identified and described in the two categories *Effortless method for pain assessment* and *Positive influences on the care*. The professional caregivers' experiences of the patients' use of the pain assessment method resulted in the categories *Shifting outlook towards the pain assessment method* and *Positive patient influences*, emerging from seven sub-categories.

One selected original quotation, translated into English, from each category is presented to illuminate the experiences of the patients and the professional caregivers.

a) ***Effortless method for pain assessment.*** The pain assessment method was regarded as being effortless in spite of the patients' state of health. The patients made it clear that they could handle the digital pen, which was looked upon and used as an ordinary pen, and they could easily interpret all parts of the pain diary. They did not need help from their next-of-kin in writing with the digital pen. The spouses did not help with writing but could help with other things, such as reminding the patient to perform the assessment. It was furthermore revealed that understanding of the technology and pain assessment system was limited but the motivation to use the method was not decreased by technology problems.

“No, I was never too bad to use the pen.” (*Patient 4*)

b) ***Positive influences on the care.*** The patients perceived an increased and improved contact with their professional caregivers. Using the digital pen and pain diary was described as superior to using the phone for reporting the pain, since the former was seen as a less intrusive way. The method also made it possible to report improvements in pain in an easy way. The patients took a greater part in their own care. By keeping the original pain diaries the patients could go back and see for themselves how the pain had varied over time and this led to a wish to continue with the pain assessment after the period using the digital pen. What the patients appreciated above all with the pain assessment method was that it resulted in a sense of increased security.

“Well I must say it's *absolutely clear* that it is positive ... and only has advantages, for it is considerably faster. I can reach them [the caregivers] easier and that means *a feeling of greater peace for me*, which is the main thing.” (*Patient 12*)

c) ***Shifting outlook towards the pain assessment method.*** The professional caregivers' outlook was initially cautious due to low expectations concerning the patients' abilities and due to uncertainty about how to use the system. Contemporary organisational changes put the caregivers under stress and the pain

assessment system was described as increasing their workload. Although they were reluctant to use the system and change their way-of-working, the caregivers experienced positive outcomes. They found it valuable to receive information on the patients' pain situations and they learned to be more aware of and focus on the patients' pain. The improvement suggestions for future use comprised assessment of all pain locations and assessment of more symptoms.

"Somehow we use the VAS in a little more objective way than we did before ... I don't say in words that the patient is in *more* or *less* pain, instead I say that 'The patient assesses pain VAS six'. Later on in the afternoon he perhaps assesses VAS three ... then it is not *my* subjective judgment of what he tells me, but *his* objective." (Nurse 1)

- d) **Positive patient influences.** This category comprised the professional caregivers' perception of the participating patients' experiences. The caregivers saw that the patients had benefited from the pain assessment method by means of increased participation in their care, increased security and by improved responses to pain fluctuations in terms of changes in treatment. According to the caregivers, patients sometimes seemed to be somewhat amused by using the pain assessment method, which led to continued pain assessment after the period.

"I believe that some patients have received much quicker help than if we had handled this in routine care." (Physician 2)

3. Discussion

Digital pen technology was found suitable in regard to the fact that the HBHC clinic used paper-based medical records. Of the interviewed patients only one was a woman while of the interviewed spouses all four were women, which can be seen as a limitation to the study. However, palliative patients at their end-of-life constitute a very frail group, which means that it is not always possible to create similar and complete cases [18]. The qualitative descriptive and explorative design for evaluation comprising a content analysis was used in order to capture patients' and professional caregivers' experiences of the pain assessment method. Qualitative content analysis was deemed suitable due to the fact that it focuses on the subject and context, and it points out differences and similarities between and within categories. Furthermore, it preserves the core of the original analysis material [19].

During the study the participating patients assessed their momentary pain three times per day at regular times. The advantages of letting patients assess their momentary pain several times per day, instead of asking the patients to recall and summarize their pain over a period of time (e.g. one day), have been shown by, for example, Stone et al. [20] and Lefebvre and Keefe [21].

The success of home telehealth depends on both patients' and professional caregivers' acceptances. Home telehealth interventions change the caregivers' practice patterns and have an impact on their way-of-working. Also, the level of patient acceptability is influenced by the way in which patients understand the technology and intervention [22]. In a Canadian home telehealth project, involving palliative patients using video visits, the caregivers were concerned by the impact of the new technology on their usual work patterns; time and effort needed for the technology were seen as a

large investment, whereas the benefits were less obvious [23]. The professional caregivers in our study reported reactions which showed initial similarity with the Canadian study; the caregivers at HBHC showed an initial cautious outlook due to their own reluctance to use the system and change their way-of-working. Contemporary organisational changes put the HBHC caregivers under stress and the pain assessment system was described as contributing to an increase in the workload, since they had to use the computer-based system and handle both the equipment and all pain assessments on paper. Due to low expectations of the patients' abilities to use the pain assessment method certain patients, especially older women, were never asked to participate. This can be coupled to the professional caregivers' own feelings of uncertainty towards the technology which were transferred to the patients. To compensate for possible caregiver reluctance towards the technology, individual positive experiences of technology usage were important. Such positive experiences were described and contributed to a shifting outlook towards the pain assessment method. Another positive experience expressed was the fact that they had become more aware of the patients' pain and that also the patients had become more aware, which could make pain assessment and reporting easier since they "spoke the same language" by using VAS as an instrument.

The patients in the Canadian study [23] showed quite similar reactions to the patients in our study; the Canadian patients had positive reactions towards the technology though they were not familiar with it, also, they seemed "readier" than their caregivers to use the monitoring equipment. The patients in our study found the pain assessment method as being effortless in spite of severe illness and limited understanding of the technology and system intervention, and they did not need help from their next-of-kin in performing the assessments. Furthermore, the patients expressed that they took a greater part in their own care during the pain assessment period. Their motivation was that it felt good to co-operate with the professional caregivers and to see one's own pain variations and have control over consumed extra doses of analgesics.

The professional caregivers experienced several positive patient influences from the pain assessment method, which were in accordance with the results from patients'/spouses' interviews and patients' ease-of-use questionnaire and which contributed to the shifting outlook, namely that the patients took a greater part in their own care and that they experienced an improved contact with their caregivers, which in turn led to a sense of increased security. There was also a quick response to variations in the patients' health status by means of changes in medical treatment, as shown in the medical records.

For palliative patients in advanced home healthcare there is a need for ease-of-use interfaces in handling technology. Therefore digital pen technology is suitable for the assessment of symptoms since these patients often have a limited capacity to handle technology due to their state of health. By using pain diary forms on digital paper in combination with digital pen and wireless Internet technologies, our system combines many of the good parts from systems using electronic diaries and systems using traditional pen and paper-based diaries.

4. References

- [1] Dudgeon DJ, Harlos M, Clinch JJ. The Edmonton Symptom Assessment Scale (ESAS) as an audit tool. *J Palliat Care*; (1999) Autumn;15(3):14-9.
- [2] Strömngren AS, Groenvold M, Pedersen L, Olsen AK, Spile M, Sjogren P. Does the medical record cover the symptoms experienced by cancer patients receiving palliative care? A comparison of the record and patient self-rating. *J Pain Symptom Manage*. (2001) Mar;21(3):189-96.
- [3] Davies E and Higginson IJ (Editors) (2004) *The Solid Facts—Palliative care*. ISBN 92890 10916. Available from <http://www.euro.who.int/> (accessed 2007-10-28).
- [4] Grossman SA. Assessment of cancer pain: a continuous challenge. *Support Care Cancer*. (1994) Mar; 2(2):105-10.
- [5] *Cancer Facts & Figures 2006*. The American Cancer Society, ACS (Internet homepage). Copyright 2006 American Cancer Society. Available from <http://www.cancer.org> (accessed 2007-10-28).
- [6] Stone AA, Broderick JE, Schwartz JE, Shiffman S, Litcher-Kelly L, Calvanese P. Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. *Pain*. (2003) Jul;104(1-2):343-51.
- [7] de Wit R, van Dam F, Hanneman M, Zandbelt L, van Buuren A, van der Heijden K, Leenhouts G, Loonstra S, Abu-Saad HH. Evaluation of the use of a pain diary in chronic cancer pain patients at home. *Pain*. (1999) Jan;79(1):89-99.
- [8] Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. *Control Clin Trials*. (2003) Apr;24(2):182-99.
- [9] Jamison RN, Raymond SA, Levine JG, Slawsby EA, Nedeljkovic SS, Katz NP. Electronic diaries for monitoring chronic pain: 1-year validation study. *Pain*. (2001) Apr;91(3):277-85.
- [10] Gaertner J, Elsner F, Pollmann-Dahmen K, Radbruch L, Sabatowski R. Electronic pain diary: a randomized crossover study. *J Pain Symptom Manage*. (2004) Sep; 28(3):259-67.
- [11] Bird J. Assessing pain in older people. *Nurs Stand*. (2005) Jan 19-25; 19(19):45-52; quiz 53-4. Review.
- [12] LeeCF, KuoCC. Difficulties on Small-Touch-Screens for Various Ages, in *Universal Access in Human Computer Interaction. Coping with Diversity* (2007) Vol 4554, p 968-974. Springer Berlin / Heidelberg
- [13] Cole E, Pisano ED, Clary GJ, Zeng D, Koomen M, Kuzmiak CM, Seo BK, Lee Y, Pavic D. A comparative study of mobile electronic data entry systems for clinical trials data collection. *Int J Med Inform*. (2006); 75:722-9..
- [14] Despont-Gros C, Landau R, Rutschmann O, Simon J, Lovis C. The digital pen and paper. Evaluation and acceptance of a new data acquisition device in clinical settings. *Methods Inf Med* (2005) 44(3):359-68.
- [15] Dykes PC, Benoit A, Chang F, Gallagher J, Li Q, Spurr C, McGrath EJ, Kilroy SM, Prater M. The feasibility of digital pen and paper technology for vital sign data capture in acute care settings. *AMIA Annu Symp Proc*. (2006) 229-33.
- [16] Lind L, Karlsson D. A system for symptom assessment in advanced palliative home healthcare using digital pens. *Med Inform Internet Med*. 2004 Sep-Dec;29(3-4):199-210.
- [17] Lind L, Karlsson D, Fridlund B. Patients' use of digital pens for pain assessment in advanced palliative home healthcare. *Int J Med Inform*. 2008 Feb;77(2):129-36. Epub 2007 Mar 23.
- [18] Caraceni A, Cherny N, Fainsinger R, Kaasa S, Poulain P, Radbruch L, De Conno F. Pain measurement tools and methods in clinical research in palliative care: recommendations of an Expert Working Group of the European Association of Palliative Care. *J Pain Symptom Manage*. (2002) Mar;23(3):239-55.
- [19] Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Educ Today*. (2004) Feb;24(2):105-12.
- [20] Stone AA, Schwartz JE, Broderick JE, Shiffman SS. Variability of Momentary Pain Predicts Recall of Weekly Pain: A Consequence of the Peak (or Salience) Memory Heuristic. *Pers Soc Psychol Bull*. (2005) Oct; 31(10):1340-6.
- [21] Lefebvre JC, Keefe FJ. Memory for pain: the relationship of pain catastrophizing to the recall of daily rheumatoid arthritis pain. *Clin J Pain*. (2002) Jan-Feb; 18(1):56-63.
- [22] Demiris G, Shigaki CL, Schopp LH. An evaluation framework for a rural home-based telerehabilitation network. *J Med Syst*. (2005) Dec; 29(6):595-603.
- [23] Gagnon MP, Lamothe L, Hebert M, Chanliau J, Fortin JP. Telehomecare for vulnerable populations: The evaluation of new models of care. *Telemed J E Health*. (2006) Jun;12(3):324-31.

Reusing Models of Actors and Services in Smart Homecare to Improve Sustainability

Ståle WALDERHAUG^{a+b}, Erlend STAV^a and Marius MIKALSEN^a,

^a SINTEF ICT, Trondheim, Norway

^b Department of Computer Science, University of Tromsø, Norway

Abstract: Industrial countries are faced with a growing elderly population. Homecare systems with assistive smart house technology enable elderly to live independently at home. Development of such smart home care systems is complex and expensive and there is no common reference model that can facilitate service reuse. This paper proposes reusable actor and service models based on a model-driven development process where end user organizations and domain healthcare experts from four European countries have been involved. The models, specified using UML can be reused actively as assets in the system design and development process and can reduce development costs, and improve interoperability and sustainability of systems. The models are being evaluated in the European IST project MPOWER.

Keywords: Homecare, Smart house, Service Oriented Architecture, System architecture, UML, Model-driven software development, standardisation, HL7

Introduction

Home care is a concept where new technological solutions allow the elderly to live independently at home. The consumers, typically elderly, chronically ill and cognitive disabled are empowered by state-of the art information and technology in their homes to achieve the overall goal of aging in place[1-3].

The technological advances are being deployed into the homecare domain, and many projects are working on *smart house* technology [4, 5] for homecare. This concept needs, in addition to technical devices, to be supported by a team of actors including family members, healthcare personnel and non-healthcare services organizations. A smart homecare system is both complex and expensive to build. Efficient development of complex systems should seek to *reuse* components and services through abstraction to “*manage complexity and guarantee continuity*”[6]. Krueger states that “*in order to reuse artefacts, software developers must either be familiar with the abstractions a priori or must take time to study and understand the abstractions.*”[7]. The importance of understanding the domain concepts is emphasized by Beyer et al in [8] where they state that “*to reduce the effort for system evolution it is highly desirable to incorporate generic components, that can be reused in different contexts*”.

Lenz, Beyer and Kuhn elaborate on this in [9] and argues for a separation of domain concepts and system implementation: *“in order to cope with domain evolution, modelling of domain concepts should be separated from IT system implementation. IT systems should be implemented by IT experts and medical knowledge should be modelled and maintained by domain experts.”* A reusable reference model for smart homecare can form the basis for more efficient development of smart homecare systems and reuse of its services[8, 10-12]. This article presents the work done in the MPOWER project [13] where a model of the core business process actors and services in smart homecare systems has been specified. Two research questions are addressed:

- Which actors (persons and systems) are involved in the teamwork treatment of smart homecare service consumers?
- Which information services are needed to support the (treatment) processes?

The focus of the work is on actors and information service support for elderly and cognitive disabled. The project result is a formal representation of actors and services that enable service reuse and increase the understanding of actor-service dependencies.

The remainder of the paper is organized as follows: First the methods and materials for specifying the actor and service models are described. Next, the core of these models is presented before the implications they may have on the development of sustainable healthcare systems are discussed.

1. Methods and Materials

Domain experts and user groups for elderly and cognitive disabled in four European countries were involved in the requirements process; Austria, Poland, the Netherlands and Norway [14]. The process produced the “User Scenario Specification” describing the problems experienced by the target groups and the planned assistive smart homecare services. Figure 1 shows the overall iterative approach for the work.

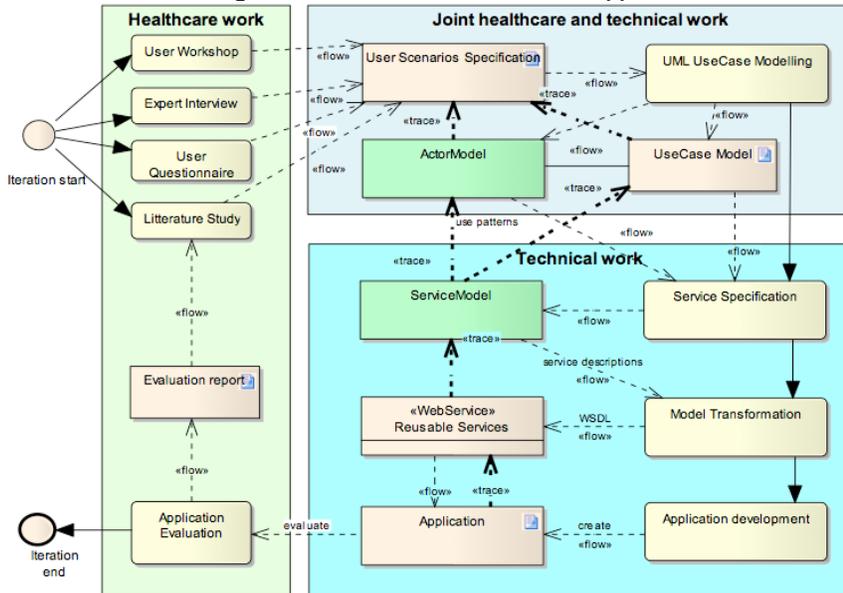


Figure 1: The iterative model-driven development process used to identify actors and services. Artefacts are shown as rectangles whereas the activities are denoted as rectangles with rounded corners.

This user requirements document was used as the basis for UML [15] UseCase Modelling activity that produced the UML Actor and UseCase Models. The modelling work was a joint effort by healthcare personnel and system architects, as recommended by Lenz et al in [9]. The Service Specification activity uses the Actor and UseCase models as input to create the Service Model which is used in the Model Transformation [16-18] activity for creating Reusable Services (Web Services). Finally, the Reusable Services are used by application developers to implement applications that will be evaluated by healthcare personnel. The evaluation report (from Application Evaluation activity) is used as input for the succeeding iteration. The bold arrows in Figure 1 highlight traceability links between artefacts [19].

The Use Case Model, Actor Model and Service Model were developed using Unified Modelling Language (UML) with the IBM Profiles for Software Services [20] according to Service Oriented Architecture (SOA) [21] concepts. The services in the Service Model were identified according to best practice SOA in general [21], and for healthcare especially [22, 23].

2. Results

In this paper, only a selection of the most important elements from these models is presented in detail. Full specifications are accessible on the MPOWER Website [13]. The use case and service modelling activities resulted in three UML models: the Actor Model, the Service Model, and the Use Case model that relates the two first.

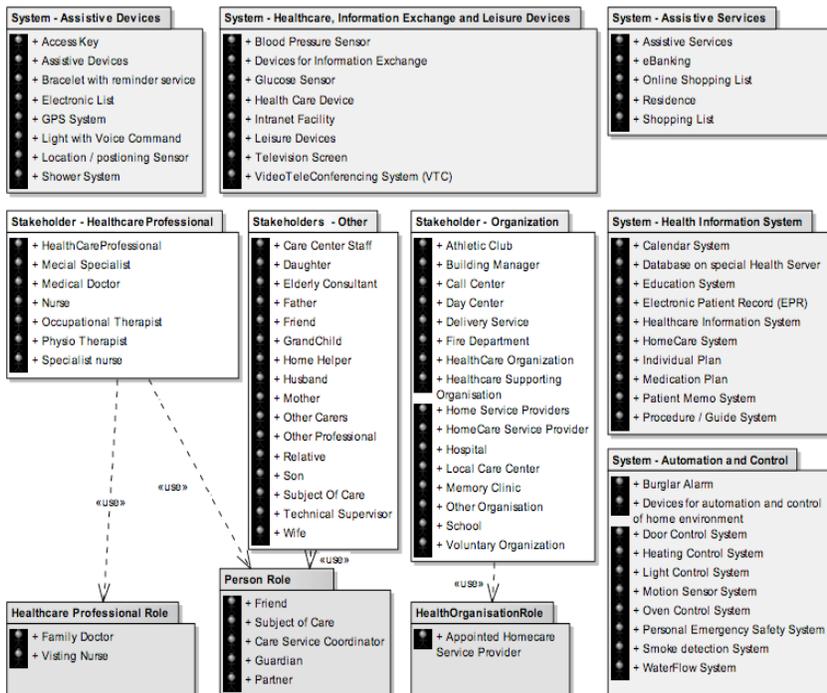


Figure 2: The ActorModel showing the elements of the system, stakeholder and role parts.

The Actor Model (Figure 2) has three main parts: System (light grey), Stakeholders (white) and Roles (dark grey). The stakeholders can have different roles as shown in Figure 3. The roles that the stakeholders can take are modelled as a dependency link, e.g., only a Nurse or a Specialist Nurse can have the role as a Visiting Nurse. All Healthcare Professionals can be a patient themselves (role Subject of Care).

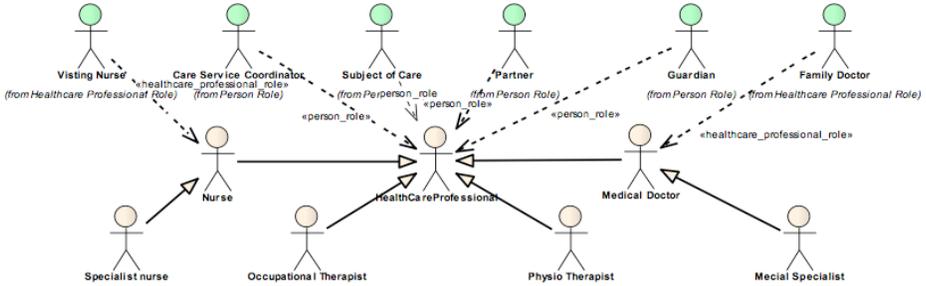


Figure 3: The relationships between healthcare professional actors and roles as defined in the ActorModel. The UseCase model defines activities that the actors (and roles) participate in. These activities are the link to the services in the ServiceModel. Figure 4 shows a use case diagram for calendar management activities involving systems, stakeholders and roles.

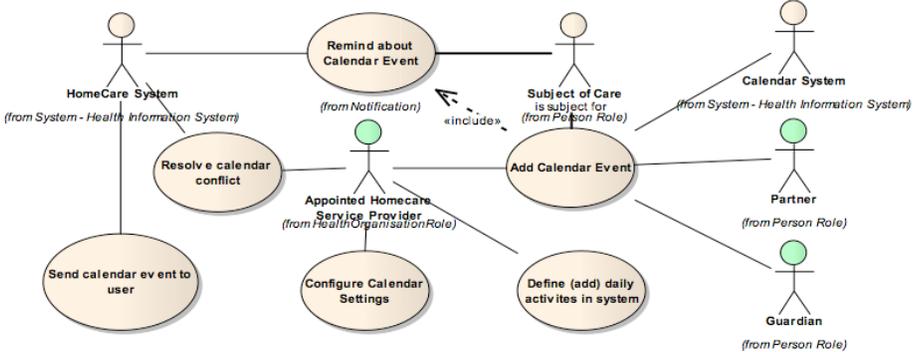


Figure 4: Actors, Roles and Use cases in the usecase diagram for Calendar Management services

From the scenarios and use cases five categories of services were specified (Figure 5) using the service identification principles described in [23].

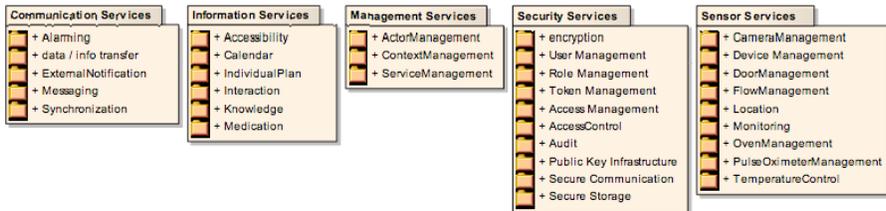


Figure 5: The service categories in the ServiceModel. The elements of each category are reusable services

3. Discussion

The Actor Model specified from the User Services Specification includes formal actor and role specifications in UML that allows for reuse across functional domains such as

medication management and home automation services. A common and formally specified Actor Model means that 1) Actor-System (service) interaction can be precisely specified, 2) access control for services can be derived from use cases with actor interaction, 3) Actor-System dependencies can be traced through trace links, and 4) Systems and services can be compared to each other.

The elements of the Actor Model were agreed by all four countries participating in the specification process, and it is conform to the CEN TC251 CONTSYS [24] standard and compatible with those actors described in [25] and [5]. An important finding when modelling the Actor Model was the feasibility of using *roles* in use case modelling. In many cases, it is not the *actor* itself that interacts with the system, rather an actor taking a certain role. Using this concept, the constraints will be put on which *actor* that can take a *role* instead of modelling use cases for different actors in different contexts. Figure 4 shows an example where the *Appointed Homecare Service Provider*, *Partner* and *Guardian* roles are used to interact with the add calendar event activity.

The services identified for the Service Model are being implemented and used in two proof-of-concept applications in the MPOWER project [13]. The applications communicate with the services through HL7 messages, and all underlying complexity is handled by the services. The results from these developments will evaluate the services' reusability in the domain. Preliminary development results show that:

- Applications can be developed more rapidly by reusing high-quality services [8].
- Functionality can be reused across applications and organisations and nations, e.g., sending SMS, PKI, calendar management, medication list management.
- Clearly defined actors will improve the validity of the system services being developed and improve sustainability[6, 26]
- The gap between business processes and supporting information systems can be shortened by applying the Service Oriented Architecture concepts [10, 21-23]
- Aligning service and actor descriptions with national and international standards will promote standardisation and facilitate reuse of services and components across organisations and nations, thus improving interoperability[27]

Reuse of software and design is not trivial. Krüger states that "*for a software reuse technique to be effective, it must reduce the cognitive distance between the initial concept of a system and its final executable implementation.*" The Actor and Service models presented herein are the results of a formal process to reduce this distance.

The results presented herein are generalisations, and may not be directly applicable to all domains without prior local adaptations. The actor and service models can serve as reference models from which nation and organization specific models can be developed in accordance with the prevailing ways of organizing care and legislations. The specification of reference models must be supported by standards developing organisations such as CEN TC251 and HL7 [28]. The proposed models being implemented and evaluated in the MPOWER Project [13], and will be presented for the HSSP project [29] and national standardisation bodies in Europe.

4. References

- [1] Magnusson, L.H., E.; Borg, M, *A literature review study of information and communication technology as a support for frail older people living at home and their family carers*. Technology and disability, 2004. **16**(4).
- [2] Wacanta, J.e.a., *Number of dementia sufferers in Europe between the years 2000 and 2050*. European Psychiatry, 2003. **18**: p. 306-313.

- [3] Demiris, G., et al., *Older adults attitudes towards and perceptions of smart home technologies: a pilot study*. Medical Informatics and the Internet in Medicine, 2004. **29**: p. 87-94.
- [4] Helal, S.C., *The Gator Tech Smart House: A Programmable Pervasive Space*. Computer, 2005. **38**(3): p. 50.
- [5] Stefanov, D.H., Z. Bien, and W.-C. Bang, *The smart house for older persons and persons with physical disabilities: structure, technology arrangements, and perspectives*. IEEE transactions on neural systems and rehabilitation engineering, 2004. **12**(2): p. 228-50.
- [6] Shaw, M., *Abstraction techniques in modern programming languages*. IEEE Software, 1984. **1**(4): p. 10-26.
- [7] Krueger, C.W., *Software reuse*. ACM Computing Surveys (CSUR), 1992. **24**(2): p. 131-183.
- [8] Beyer, M., et al., *Towards a flexible, process-oriented IT architecture for an integrated healthcare network*, in *Proceedings of the 2004 ACM symposium on Applied computing*. 2004, ACM Press: Nicosia, Cyprus.
- [9] Lenz, R., M. Beyer, and K.A. Kuhn, *Semantic integration in healthcare networks*. International journal of medical informatics, 2007. **76**(2-3): p. 201-207.
- [10] Kawamoto, K. and D.F. Lobach, *Proposal for Fulfilling Strategic Objectives of the U.S. Roadmap for National Action on Decision Support through a Service-oriented Architecture Leveraging HL7 Services*. Journal of the American Medical Informatics Association, 2007. **14**(2): p. 146-155.
- [11] Lenz, R., M. Beyer, and K.A. Kuhn, *Semantic integration in healthcare networks*. International journal of medical informatics, 2006.
- [12] Lenz, R. and K.A. Kuhn, *Towards a continuous evolution and adaptation of information systems in healthcare*. International journal of medical informatics, 2004. **73**(1): p. 75-89.
- [13] The MPOWER Consortium. *MPOWER Website*. 2007 [cited 2007 June 15]; Available from: <http://www.mpower-project.eu>.
- [14] Prazak, B., et al. *User Requirements as Crucial Determinants for the Development of New Technological Solutions for Elderly Care - Exemplified in an European Project*. in *AAATE*. 2007. San Sebastian, Spain: IOSPress
- [15] Object Management Group (OMG), UML 2.0 Superstructure FTF Rose model containing the UML 2 metamodel, Object Management Group (OMG), 2005, <http://www.omg.org/cgi-bin/apps/doc?ptc/04-10-05.zip>
- [16] Frankel, D., *The MDA Journal: Model Driven Architecture Straight from the Masters*. 1st ed. 2004, Tampa, Florida, USA: Meghan-Kiffer Press. 219.
- [17] Mellor, S.J., *MDA Distilled: Principles of Model-Driven Architecture*. 2004.
- [18] Miller, J. and J. Mukerji, MDA Guide Version 1.0.1, Object Management Group (OMG), 2003-06-13, 2003, <http://www.omg.org/docs/omg/03-06-01.pdf>
- [19] Walderhaug, S., et al., *Traceability in Model-Driven Software Development*, in *Designing Software-Intensive Systems: Methods and Principles*, P. Tiako, Editor. 2008, IDEA Group: Langston.
- [20] Johnston, S. *UML 2.0 Profile for Software Services*. 2005 [cited 2007 June 15]; Available from: http://www.ibm.com/developerworks/rational/library/05/419_soa/.
- [21] Erl, T., *Service-Oriented Architecture Concepts, Technology, and Design*. The Prentice Hall Service-Oriented Computing Series ed. T. Erl. 2006, Crawfordsville, Indiana, USA: Prentice Hall.
- [22] Honey, A., et al., SOA4HL7 Architecture Document, Health Level Seven, August 16, 2006,
- [23] Honey, A. and B. Lund, Service Oriented Architecture and HL7 v3: Methodology, HL7 Service Oriented Architecture Special Interest Group (SOA SIG), November 10, 2006, 2006,
- [24] CEN TC251, EN 13940-1: Health Informatics - System of Concepts to Support Continuity of Care - Part 1: Basic Concepts, European Committee for Standardization, September 26, 2006,
- [25] Toivanen, M., et al., *Gathering, Structuring and Describing Information Needs in Home Care: A Method for Requirements Exploration in a "Gray Area"*. MEDINFO 2004: Building High Performance Health Care Organizations, 2004: p. 7-11.
- [26] Coiera, E. and E.J. Shovenga, *Building a Sustainable Health System*, in *IMIA Yearbook of Medical Informatics 2007: Biomedical Informatics for Sustainable Health Systems*, A. Geissbuhler, R. Haux, and C. Kulikowski, Editors. 2007, Schattauer Publishers. p. 250.
- [27] Mykkaenen, J., *Specification of Resuable Integration Solutions in Health Information Systems*. Department of Computer Science. Vol. Doctoral Dissertation. 2007, Kuopio, Finland: Kuopio University Publications, Finland 133.
- [28] Blobel, B., *Authorisation and access control for electronic health record systems*. International Journal of Medical Informatics, 2004. **73**(3): p. 251-257.
- [29] HSSP Project, The HSSP Roadmap: HSSP, Version 1.0, Joint HL7-OMG Healthcare Services Specification Project, October 10, 2007, 2007,

A System for Monitoring Physical Activity Data Among People with Type 2 Diabetes

Eirik ÅRSAND^{a,b,1}, Odd-Arne OLSEN^a, Ragnhild VARMEDAL^a,
Willy MORTENSEN^c and Gunnar HARTVIGSEN^{a,b}

^a Norwegian Centre for Telemedicine, University Hospital of North Norway, Norway

^b Department of Computer Science, University of Tromsø, Norway

^c SensoTek, Tromsø, Norway

Abstract. Trends towards lower levels of physical activity have raised health concerns. Tools to capture, store and use information about physical activity might improve motivation to increase the level of such activity. This is especially important for Type 2 diabetes, since physical activity is one of the key components in achieving healthy blood glucose values. Over a period of four months, 15 people with Type 2 diabetes provided us with input on how a mobile system needs to be put together. Generally, they answered that such tools must be integrated as well as possible with their other daily tools and clothing. Based on their inputs, we built a sensor system for monitoring physical activity. The system automatically and wirelessly reports the accumulated number of steps taken, using a mobile phone as the patient terminal. We asked 1001 persons about their use of step counters/pedometers. About 6.5 % of them use such a device daily and about 20 % daily, weekly or monthly. Our concept differs from others of this nature in its simplicity, size and integration with other relevant patient data. It is fully manageable by patients themselves as a self-help tool.

Keywords. Patient monitoring, Data acquisition-data capture, Human interfaces.

Introduction

According to the World Health Organisation, physical inactivity and unhealthy diets are among the leading causes of the major noncommunicable diseases, including Type 2 diabetes, and contribute substantially to the global burden of disease, death and disability [1]. Studies have clearly indicated that increased physical activity both reduces the risk of developing Type 2 diabetes and is positive for those already diagnosed with the disease [2]. Growing awareness of the increase in obesity and Type 2 diabetes as well as a general concern with health and fitness have resulted in a strong focus on the use of step counters/pedometers as a tool for self-monitoring of physical activity. This article focuses on secondary prevention for people with Type 2 diabetes. The system presented may also hold potential for primary prevention, i.e. to be used by healthy people who want to reduce their chances of developing health problems.

¹ Corresponding Author: Eirik Årsand, Norwegian Centre for Telemedicine, University Hospital of North Norway, Sykehusvn. 23, N-9019 Tromsø, Norway; E-mail: Eirik.Arsand@telemed.no.

Aiming for Everyday Tools

Advances in biomedical engineering have brought us a number of new diagnostic tests and disease management tools. Information and communication technology have provided us with hardware and software which offer great benefits for health care applications. New wireless communication standards and the miniaturisation of electronic components have made it possible to make both wearable patient sensors and handy patient terminals. Mobile phones have grown into small computers in functionalities, often referred to as “Pocket PCs” or “Smartphones”, and are used as the patient terminal in our approach. By utilising and building on the three main components: sensor, wireless communication and patient terminal, we aim to design an application for automatic gathering of the user’s physical activity, and to provide a beneficial use of this data in an easy way. Depending on the patient’s needs and health challenges, it must be easy to adjust the system to automatically transfer the level of physical activity each hour, every two hours, twice a day, daily, or at other intervals. This calls for a sensor application with high usability, light in weight, small in size, and with a battery life of at least 6 months. When properly attached to the user it should be unnoticeable during daily activities. Vital factors therefore include miniaturisation of the electronics, minimal power consumption, as few buttons as possible and automatic and wireless data transfer. All the functionalities and user adjustments should be controlled from the patient terminal. The application presented is part of the wearable eHealth system [3], i.e. the “Easy Health Diary”, which gathers data on blood glucose values, nutrition habits, and the physical activities as described here.

State-of-the-Art for Non-Obtrusive Registration of Physical Activity

Step counters and pedometers are usually referred to as the same physical device, where the difference is only in functionality. Pedometers also calculate the distance walked, but this requires the user to enter the stride length. Today, step counters are mainly attached to the belt on the hip and have a built-in LCD to display the number of steps taken. One also finds MP3 players and mobile phones that include step counters as a fitness feature. Other ways of measuring physical activity include using heart rate monitors or global positioning system (GPS) devices. Heart rate monitors have the disadvantage that the user has to wear a transmitter chest belt, something most people find too obtrusive to wear on a daily basis. GPS is a good tool for measuring one’s movement, but is mainly useless indoors due to lack of signal strength from its satellites, and also needs frequent recharging due to high power consumption.

It has not yet been possible to find a totally unobtrusive system for easily attaching physical activity measurement sensors to people. Examples of innovative systems are the SensVest [4], which requires the user to wear a specific vest with embedded sensors; the MPTrain system [5], which uses music to improve exercise performance and is a combined heart-rate monitor and movement monitor device; the mobile-phone-sized NEAT-o-games application [6] using Bluetooth for transfer of the accelerometer data to a PDA; and the multi-modal sensor board and Bluetooth unit enclosed in a box worn on the waist for wireless monitoring [7]. Common to all of these is that they have a short battery life-length, are too large to be worn for long periods and/or require substantial user interaction for achieving the intended functionalities.

1. Methods

1.1. User-Centred Design Methods

As part of our design process, 15 people with Type 2 diabetes were engaged over a period of four months. The participants were recruited through letters sent to all members of the local diabetes association that were between the ages of 40 and 70 years. They gave feedback both on how they would like a comprehensive mobile-phone-based self-help system to be put together as a sustainable tool [8], and specifically on the physical activity parameter. The methods used were focus groups, paper prototyping and sketching, prototyping, questionnaires, and giving the informants small amounts of homework between the five focus group meetings. We gave the users an off-the-shelf step counter for use between the two first meetings and asked them to report their experiences. For reference on step counter use, we prepared four questions on use of health monitoring devices as part of the “eHealth Trends survey” in April 2007, with 1001 informants using telephone interviews (CATI).

1.2. Microcontroller Design Methods

The step-counter device was based on a microcontroller to provide the intelligence needed for performing logical operations. Requirements for the microcontroller included small size, low power consumption, flash memory, low cost, and I/O drivers with high flexibility. The Atmel ATmega164P 8-bit RISC microcontroller was found to fulfil these and became the sensor system’s main building block. To program the step-counter application features, special software and hardware programming tools were needed: the Integrated Development Environment (IDE) “AVR Studio” and the “AVR STK500” starter kit. The IDE provided an assembler, an editor, and a simulator, enabling possibilities for testing the functionalities during the development process. In addition, the emulator “Atmel AVR JTAGICE mkII” was used to test and edit the step-counter software on the actual microcontroller, in circuit. To set up communication to the wireless interface, the microcontroller’s asynchronous serial interface was used.

1.3. Data Transfer Methods and Electronics Design Methods

Bluetooth is at present the only short-range communication protocol which is fully implemented as a short-range communication protocol in mobile phones, and was therefore chosen for communicating the data to the patient terminal. The OEMSPA311i-04 Bluetooth adapter, 16 mm by 36 mm class 1, from connectBlue AB was used. An asynchronous serial databus without handshaking (RS-232-like protocol) was used for data transfer. The module was configured so that when powered up it attempts to connect to the patient terminal, informing the microcontroller when contact is achieved, and finally transmits the data received by its UART. At the patient terminal, an application was made to constantly listen for incoming data, storing it on the phone and sending it to the server and/or presenting it to the user on the mobile phone screen. The application is implemented in C# using .NET Compact Framework, and runs on the MS Windows Mobile 6.0 platform. To initiate transfer of step count data without requiring user interaction, we enabled automatic transfers at configurable intervals. The user can also transfer data on demand by pushing the only button on the step counter. This button was recessed level with the chassis to avoid accidental data transfer.

A mechanical sensor was chosen to minimise power consumption and microcontroller programming efforts, compared with a triaxial accelerometer that would have provided more flexibility. A mechanical sensor generates ripple and needs a low-pass filter for removing unwanted signals. Using passive components for the filter in addition to software routines made it possible to distinguish steps from noise in the signal. The sensor application's power consumption ranges from $2\mu\text{A}$ to 40mA and requires a battery bigger than button-cell batteries: the Saft LS14250. The main current consumer, the Bluetooth adapter, has a built-in sleep mode, but even in this mode it consumes more than 1mA . This problem was solved using an external power control. The drawback is a longer start-up time for the Bluetooth module, but is acceptable since the result is an average power-up to power-down time of just 3 seconds.

Since the sensor and the mobile phone are paired, a link key valid only between these two Bluetooth units is created, providing security considered sufficient for this kind of data. The electronics are encapsulated in a plastic chassis, enabling an internal antenna and only a slight loss in signal strength. It was decided to have no LCD on the step counter, since we wanted all interaction to be controlled from the patient terminal.

2. Results

2.1. Users' Feedback

In the 2007 survey on eHealth trends [9], we included a question about how often the informants in the Norwegian sample had used a step counter/pedometer during the last half year. A professional polling agency conducted 1001 interviews, in which 6.6 % of informants reported daily use, 7.4 % weekly use, 6.3 % monthly use, 12.2 % more seldom than monthly use, and 67.5 % no use during the last half year; see Figure 1. The cumulative percentage of daily, weekly or monthly use was 20.3 %. In comparison, a study by Eakin et al. [10] showed a pedometer use of between 5.1 % and 18.1 % during the last 18 months dependent on geographical location. In the eHealth survey, we also asked about the use of other self-monitoring devices such as pulse, blood pressure, and blood glucose monitors. Informants reported less frequent use of these devices, with cumulative monthly percentages of respectively 11.3 %, 7.7 % and 5.7 %. These results foster the belief that the parameter "step counts" is both easy to understand and motivating for people in general to monitor. Our Type 2 diabetes cohort confirms this belief, and the 15 participants expressed a positive attitude to the use of step counters. They wanted the functions of a step counter to be as automatic and easy to use as possible, and said that tools for self-help should be integrated with their daily tools and outfits. Further user needs are published elsewhere [8].

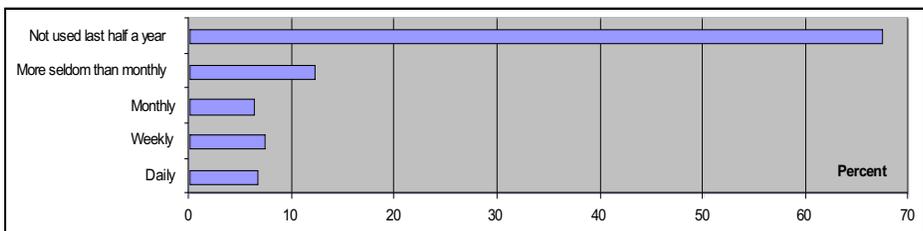


Figure 1. Frequency of step counter/pedometer use by the 1001 Norwegian informants.

2.2. The Final Prototype

The final prototype is dominated by a printed circuit board containing the Bluetooth module, the microcontroller, the movement sensor and the battery, see Figure 2. These four components are also the major contributors to the step counter's physical size of 6 x 4 x 1 cm. To filter out random movements and noise, the application will not start counting before it has received six consecutive movements within a timeframe of 8 seconds. Then it counts normally until it has not been subject to any movements for 8 seconds. The power settings have been set to -4dBm (124), providing a communication range of 10 meters. In order to optimise power saving, the embedded microcontroller is set to go to sleep between each step the user takes. A new movement event will wake the microcontroller up for registering and storing the step. Although the step-counter application so far is equipped with a movement sensor only, it is designed for adding more sensors. By embedding a temperature sensor, a light sensor, and/or a noise sensor, the system may be programmed to be context sensitive. Potential cases may be to indicate whether the user is moving indoors or outdoors, is in a dark or light environment, is alone or accompanied, thus helping the patient terminal's logic to choose the best moments for user interaction.

2.3. Test of the Step-Counter Application

At this stage the step-counter application has undergone a functional test only. During 2008, the prototype will be offered for a clinical test among 14 people with Type 2 diabetes. The functional test, performed both on a treadmill and in natural environments, showed that the step-counter application provides the user with sufficient accuracy compared with the "Omron Walking Style II HJ-113" step counter. Tests of walking at normal speed indoors and outdoors showed a difference of 2.5 % after 2 hours, 1.2 % after 7 hours and 0.9 % after 12 hours between the two step counters. The treadmill test was performed at five different walking speeds. At 1 km/h the Omron step counter had problems in registering steps at all, while our application registered all steps. At 2 km/h the difference was 4 %, at 3 km/h it was 3 %, at 5 km/h it was 7 %, and at 10 km/h there was no difference between the application and the Omron step counter.

It has also been demonstrated that the step-counter application sends data properly to the patient terminal, both automatically at preset intervals, and manually when the transfer button is pushed.

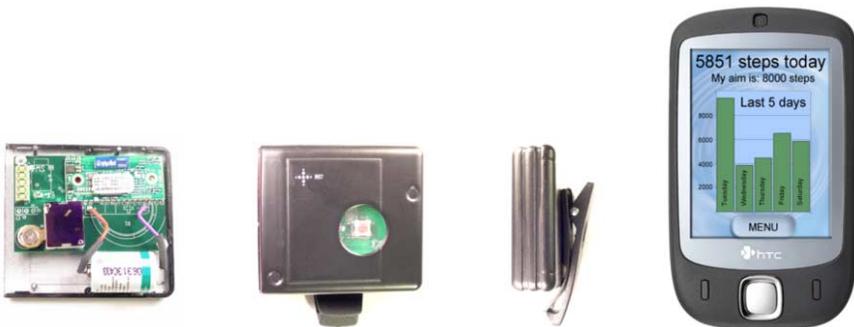


Figure 2. The final prototype; the electronics (left), front view, side view, and the patient terminal.

3. Conclusion

This step-counter application differs from others in its simplicity and the graphical feedback automatically generated on the patient terminal. Historical and current data is presented, configured to the user's own target. The system consists of two devices: the sensor to attach the belt or similar, and the patient terminal, which at the same time is the user's mobile phone.

A future goal is to reduce the sensor size to 3 x 2 x 1 cm. This probably involves miniaturisation and change of the Bluetooth adapter, the battery and the movement sensor. The trade-off between using a passive mechanical sensor and using a triaxial accelerometer will be reconsidered based on our experiences from the clinical test. We are working towards embedding more sensors for designing a context-aware patient system. The choice of microcontroller enables a transition from assembler to C, which may be necessary when making the application more advanced.

The overall goal and the main work task is to finalise a patient management system that optimises the use of physical activity data as well as blood glucose data, nutrition habit data, and specific and general patient information. By applying context awareness and self-adaptability, we hope to further increase the usability and perceived usefulness so that the concepts described become useful and sustainable patient tools.

Acknowledgement

This work was supported in part by the Research Program for Telemedicine, Helse Nord RHF, Norway and the Centre for Research-based Innovation, Tromsø Telemedicine Laboratory (TTL), Norwegian Research Council Grant No. 174934. We would like to thank our colleagues Harald Øverli Eriksen, Geir Østengen, Deede Gammon, Per-Egil Kummervold, Tove Sørensen, as well as Arne Munch-Ellingsen and Tore Syversen from our collaborative partner Telenor R&I for their valuable help.

References

- [1] World Health Organization, Global strategy on diet, physical activity and health. [Online] 2002 [cited 2007 19. October]; Available: http://www.who.int/gb/ebwha/pdf_files/WHA57/A57_R17-en.pdf.
- [2] World Health Organization, Diet, nutrition and the prevention of chronic diseases. Technical Report Series no. 916., 2003: Geneva. pp. 1-160.
- [3] E. Årsand and G. Hartvigsen, "A Wearable eHealth System for People with Type 2 Diabetes," in Scandinavian Conference on Health Informatics SHI 2005, Aalborg, Denmark, 2005, pp. 82 - 85.
- [4] J. F. Knight, et al., "The design of the SensVest," *Personal Ubiquitous Comput.*, 2005. 9(1): pp. 6-19.
- [5] N. Oliver and F. Flores-Mangas, "MPTrain: A Mobile, Music and Physiology-Based Personal Trainer," in *MobileHCI'06*, Helsinki, Finland, 2006, pp. 21-28.
- [6] Y. Fujiki, et al., "NEAT-o-Games: Ubiquitous Activity-based Gaming," in *CHI 2007*, San Jose, California, USA, 2007, pp. 2369-2374.
- [7] J. Lester, et al., "Sensing and Modeling Activities to Support Physical Fitness," in *UbiComp '05 Workshop*, Tokyo, Japan, 2005, pp. 1-4.
- [8] E. Årsand, R. Varmedal, and G. Hartvigsen, "Usability of a Mobile Self-Help Tool for People with Diabetes: the Easy Health Diary," in *IEEE CASE 2007*, Scottsdale, Arizona, USA, 2007, pp. 863-868.
- [9] H. K. Andreassen, T. Sørensen, and P. E. Kummervold, *eHealth Trends across Europe 2005-2007*. [Online] 2007 [cited 2007 8. October]; Available: http://www.telemed.no/getfile.php/551132.357.sersvdpxyyp/EHT+preliminary+report+final_5_2.pdf.
- [10] E. G. Eakin, et al., "Correlates of pedometer use: Results from a community-based physical activity intervention trial," *Int. Journal of Behavioral Nutrition and Physical Activity*, 2007. 31(3): pp. 1-7.

3. Decision Support and Knowledge Management

This page intentionally left blank

Medical Knowledge Packages and their Integration into Health-Care Information Systems and the World Wide Web

Klaus-Peter ADLASSNIG^{a,b,1} and Andrea RAPPELSBERGER^{a,b}

^a*Section on Medical Expert and Knowledge-Based Systems, Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, Austria*

^b*Medexter Healthcare GmbH², Borschkegasse 7/5, A-1090 Vienna, Austria*

Abstract. Software-based medical knowledge packages (MKPs) are packages of highly structured medical knowledge that can be integrated into various health-care information systems or the World Wide Web. They have been established to provide different forms of clinical decision support such as textual interpretation of combinations of laboratory test results, generating diagnostic hypotheses as well as confirmed and excluded diagnoses to support differential diagnosis in internal medicine, or for early identification and automatic monitoring of hospital-acquired infections. Technically, an MKP may consist of a number of inter-connected Arden Medical Logic Modules. Several MKPs have been integrated thus far into hospital, laboratory, and departmental information systems. This has resulted in useful and widely accepted software-based clinical decision support for the benefit of the patient, the physician, and the organization funding the health care system.

Keywords. Knowledge-based systems, decision support, internal medicine, laboratory medicine, infection control

Introduction

In the last few years, it has been possible to extract medical knowledge bases previously contained in the core of medical expert systems, add appropriate knowledge processing algorithms, and integrate these into various health-care information systems. These information systems might be hospital information systems (HISs), laboratory information systems (LISs), intensive-care patient data management systems (PDMSs), medical practice software systems, telemedicine applications, or health-care information or patient record systems in the World Wide Web (WWW). We name these extracted medical knowledge bases together with their corresponding knowledge processing algorithms medical knowledge packages (MKPs). Moreover, we have been

¹ Corresponding Author: Klaus-Peter Adlassnig, Section on Medical Expert and Knowledge-Based Systems, Core Unit for Medical Statistics and Informatics, Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, Austria; E-mail: klaus-peter.adlassnig@meduniwien.ac.at

² Medexter Healthcare GmbH, a spin-off of the Medical University of Vienna, Austria, acts as a vendor of medical decision support systems and services including medical content such as Arden or FuzzyArden MLMs and technical inference frameworks such as Arden or FuzzyArden rule engines (www.medexter.com).

able to render them available in several technical formats such as source code, or in compiled form, together with their corresponding carrier systems. A further option is to place these MKPs on a server and furnish them with standardized input and output structures for on-site or remote access. By means of these MKPs, the corresponding health-care information systems are equipped with knowledge-based medical intelligence.

1. Integration into Health-Care Information Systems

Integration into HIS: MKPs to support differential diagnosis in internal medicine can be integrated into the process of work-flow-oriented patient care—if necessary with additional input of medical data—and be rendered available in a structured documentation. Explanations will be provided for the eventually confirmed diagnoses, the offered diagnostic hypotheses, and the excluded diagnoses. The clinical user receives a detailed analysis of the patients' data and information about further diagnostic investigations so that each step of the diagnostic decision is traceable and remains transparent. MKPs for automated interpretation of laboratory test results can be useful on the information screen for test results. A simple query click by the user yields an interpretive text that explains the combination of laboratory test results achieved, particularly in cases of rare, complex, or even inconsistent findings.

Integration into LIS: MKPs for automated interpretation of laboratory test results can be integrated into the textual report of laboratory test results. A clinically oriented interpretive text—not only for common standard results but also for rare, complex, and unlikely test results—is automatically inserted into the report sent to the ordering physician [1]. This forms an essential quality assurance tool for the laboratory. Moreover, it saves time, avoids error, and provides clinically useful information. In the final analysis, it saves costs.

Integration into intensive-care PDMS: MKPs such as those for early identification and automated monitoring of hospital-acquired infections can be directly incorporated as alarm modules into intensive-care PDMSs. Through several data-to-symbol conversion steps, existing clinical and microbiological data are prepared in a manner that common definitions of hospital-acquired infections such as those of CDC [2], HELICS [3], or KISS [4] can be assessed for concurrence or non-concurrence in a fully automated fashion.

Integration into medical practice software systems: MKPs to support the differential diagnosis at the physician's office—possibly with a separate documentation of additional patient data—serve as add-on modules for medical practice software systems. Thus, these modules provide problem-specific documentation as well as support in terms of differential diagnosis [5].

Integration into telemedicine applications: In these systems, continuous monitoring of incoming data streams aimed at accepting normal patient data in an automated fashion and marking only those that need to be reported to a human decision maker has become increasingly important.

Integration into WWW applications: Browser-based access to MKPs permits world-wide dissemination and utilization of the medical knowledge contained in these knowledge packages [1, 6, 7]. Decision support systems based on these knowledge packages can be accessed through any WWW browser. One alternative is to make MKPs available on a server that can be directly accessed through a suitable

communication protocol. Data is passed on to data structures such as XML [8], HTTP, or HL7 [9]; the generated results are also passed on, if necessary with an explanatory report.

2. Technical Solutions

Technically, the integration of MKPs into health-care information systems and the WWW can be carried out as follows. The respective knowledge base is written in a special syntax format packed together with the knowledge processing system, and provided as a transferable file. This syntax can be processed by a compiler (or an interpreter).

One example is the standard format Arden [10–12] introduced in 1992 and now supervised by a Special Interest Group (SIG) of the HL7 committee [9]. Arden is used to represent medical knowledge in the form of Medical Logic Modules (MLMs). It was primarily intended to develop reminders, alerts, and recommendations [13–17]. The available medical knowledge corresponding to Arden syntax is compiled by an Arden compiler. However, the applicability of this syntax for formal representation of “larger” knowledge bases extends beyond the writing of simple MLMs. This has resulted in strongly inter-connected packages of MLMs which, however, can be processed without difficulty. The intended purpose is fulfilled.

We developed a complete Arden software package (see Figure 1) consisting of an ArdenServer with database, reasoning, and analysis components, an ArdenEngine, and an ArdenCompiler including an integrated Arden development and test environment (ArdenIDE, see Figure 2). Arden as a medical knowledge representation and rule-based inference system was selected because it is the only industry standard and has been further developed by HL7.

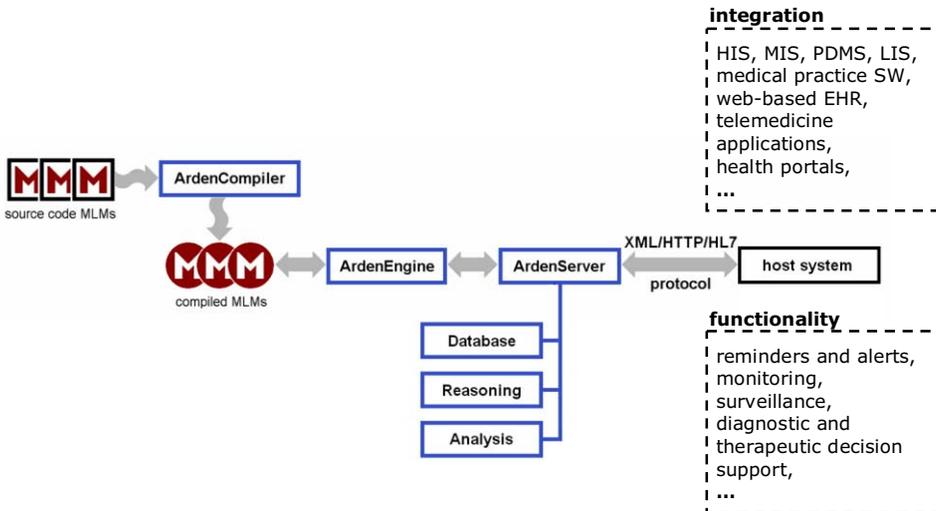


Figure 1. An Arden server that contains the respective MKPs is integrated into a host system.

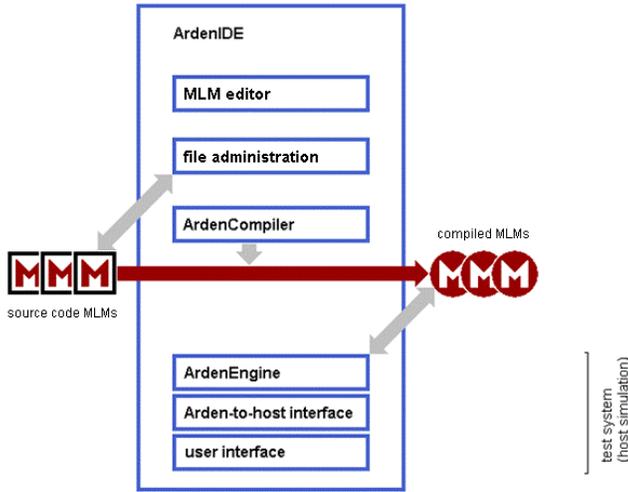


Figure 2. An ArdenIDE (integrated development environment) permits efficient development, testing, and compilation of MLMs for further processing by the ArdenEngine.

The Arden software components were themselves written in Java. They permit writing, compiling, testing, and executing Arden MLMs, which are mounted on an Arden server, controlled by an Arden engine, and connected to and triggered by the respective host health-care information system. The Arden server is equipped with several components: First, a database component to temporarily store data essential for the inference process including consecutive medical data as well as intermediate results of the inference process to allow detailed reasoning of the proposed medical decisions; second, a reasoning component to offer a line of reasoning either on a detailed level providing single data that lead to the results or on an abstracted level summarizing the detailed reasons; and, third, an analysis component to implement a mechanism for logging relevant medical or technical events for software and knowledge maintenance. The communication between the Arden server, which establishes an envelope around the Arden engine that administers the Arden MLMs, is done via data structures (XML, or HTTP, or HL7). The data structures provide access to input and output data. In this case, the so-called curly-braces problem is surmounted in such a way that the read, write, and other Arden statements with curly-braces access the Arden server's database component and not the external host health-care information system. However, we also offer—instead of the Arden server—an Arden-to-host interface. Here the curly-braces problem still exists. In some instances, data for Arden MLMs might just be transferred through the Arden “argument” and “return” statements. Our Arden engine then receives data through an XML structure and passes them on to the “argument” statements (the “return” statement runs the opposite way).

3. Present Results

At present we have a number of MKPs in Arden format which are or can be integrated into HISs, LISs, PDMSs, medical practice software systems, telemedicine applications, or the WWW.

- *Hepaxpert/Interpretation*—knowledge-based interpretation of hepatitis A, B, and C serology test results; implementations into the hospital information systems Orbis by AGFA Healthcare and Soarian by Siemens Medical Solutions were carried out successfully; Teleiatros (www.teleiatros.com) whose core consists of an Arden server provides both browser-based and direct server access to Hepaxpert/Interpretation; a remote direct server access out of Soarian was established for test purposes. The access time was a few milliseconds. Hepaxpert/Interpretation is able to interpret more than 60,000 combinations of serologic test results. We wrote a total of 7 MLMs (for each language) to carry out the respective interpretations. These contain very densely “compiled” knowledge.
- *Thyrexpert/Interpretation*—knowledge-based interpretation of thyroid hormone test results and *Toxopert/Interpretation*—knowledge-based interpretation of toxoplasmosis serology test results. Again, Teleiatros provides browser-based and direct server access to these two MKPs. Thyrexpert/Interpretation consists of 9 MLMs (per language) and Toxopert/Interpretation of 79 MLMs (per language). For both systems the MLM packages are highly interwoven and densely “compiled”.
- *RheumaDiff/Diagnosis*—medical documentation and support of decisions relating to differential diagnosis in rheumatology and *Moni/Alert-ICU*—knowledge-based early identification and automated monitoring of hospital-acquired infections at adult intensive care units. Both systems consist of packages of Arden MLMs (13 for RheumaDiff/Diagnosis per language and 47 for Moni/Alert-ICU per language). In both systems, the extended data-to-symbol conversion is also represented in MLMs. However, parts of the feature extraction in Moni still have to be done by the host health-care information system. Both are accessible through Teleiatros. Moni/Alert-ICU will shortly be integrated in a PDMS.

4. Discussion and Conclusion

MKPs are useful to support the physician in his/her medical and organizational decisions in the course of patient care. The prerequisite is that these knowledge packages should be of excellent medical quality and fully integrated into the medical work processes—whether at the clinic, the laboratory, or the physician’s office. They serve the purpose of quality assurance and potentially improve patient care, enhance the efficiency of medical work by accelerating diagnostic and therapeutic decisions, and possibly reduce the cost of care or at least render the costs transparent.

MKPs serve the purpose of clinical decision support for the individual patient by means of medical expert knowledge. The growing complexity of medical knowledge makes the use of such systems increasingly important. When used properly they reduce a large part of the attending physician’s effort in terms of repetitive mental processes

related to his specialty. They provide him time and opportunity to devote his attention to patients and contribute in great measure to the patient's safety.

The achieved results are pervading all fields of application of information systems in medicine. Initial significant steps have been taken; these are steps in the direction of knowledge-based health care for the benefit of the patient, the physician, and the organizations funding the health care system.

References

- [1] Adlassnig, K.-P. and Horak, W. (1995) Development and Retrospective Evaluation of HEPAXPERT-I: A Routinely-Used Expert System for Interpretive Analysis of Hepatitis A and B Serologic Findings. *Artificial Intelligence in Medicine* 7, 1–24.
- [2] Garner, J.S., Jarvis, W.R., Emori, T.G., Horan, T.C., and Hughes, J.M. (1996) CDC Definitions of Nosocomial Infections. In Olmsted, R.N. (Ed.) *APIC Infection Control and Applied Epidemiology: Principles and Practice*, Mosby, St. Louis, A-1–A-20.
- [3] Hospital in Europe Link for Infection Control through Surveillance (HELICS), Surveillance of Nosocomial Infections in Intensive Care Units – Protocol Version 6.1, (Based on Version 5.0 including technical amendments), September 2004, Project commissioned by the EC / DG SANCO / F/ 4, Agreement Reference number: VS/1999/5235 (99CVF4-025), 2004, pp. 1–51, <http://helics.univ-lyon1.fr/index.htm> (Accessed 15 February 2008)
- [4] Robert-Koch-Institut. <http://www.rki.de> (Accessed 15 February 2008)
- [5] Kolarz, G., Adlassnig, K.-P., and Bögl, K. (1999) RHEUMexpert: Ein Dokumentations- und Expertensystem für rheumatische Erkrankungen. *Wiener Medizinische Wochenschrift* 149, 19/20, 572–574.
- [6] Kopecky, D., Adlassnig, K.-P., Prusa, A.-R., Hayde, M., Hayashi, Y., Panzenböck, B., Rappelsberger, A., and Pollak, A. (2007) Knowledge-Based Generation of Diagnostic Hypotheses and Therapy Recommendations for Toxoplasma Infections in Pregnancy. *Medical Informatics and the Internet in Medicine* 32, 199–214.
- [7] Reitstätter, M., Vierhapper, H., Rappelsberger, A., and Adlassnig, K.-P. (2006) Quality Improvement through a Reminder System in an Outpatient Department for Thyroid Disease. In Mohammadian, M. (Ed.) *Proceedings (CD-ROM-Version) of the International Conference on Computational Intelligence for Modelling, Control & Automation (CIMCA'2006) Jointly with International Conference on Intelligent Agents, Web Technologies & Internet Commerce (IAWTIC'2006)*, IEEE Computer Society, Sydney, Australia, 27310142.pdf.
- [8] XML. <http://www.xml.org> (Accessed 15 February 2008).
- [9] Health Level Seven (HL7). <http://www.hl7.org> (Accessed 15 February 2008)
- [10] Hripcsak, G. (1994) Writing ARDEN Syntax Medical Logic Modules. *Computers in Biology and Medicine* 24, 331–363.
- [11] Peleg, M., Boxwala, A.A., Bernstam, E., Tu, S., Greenes, R.A., and Shortliffe, E.H. (2001) Sharable Representation of Clinical Guidelines in GLIF: Relationship to the Arden Syntax. *Journal of Biomedical Informatics* 34, 170–181.
- [12] Health Level Seven (2002) *Arden Syntax for Medical Logic Systems, Version 2.1*, Health Level Seven, Inc., 3300 Washtenaw Ave, Suite 227, Ann Arbor, MI 48104.
- [13] McDonald, C.J. (1976) Protocol-Based Computer Reminders, the Quality of Care and the Non-Perfectability of Man. *The New England Journal of Medicine* 295, 1351–1355.
- [14] McDonald, C.J., Hui, S.L., Smith, D.M., Tierney, W.M., Cohen, S.J., Weinberger, M., and McCabe, G.P. (1984) Reminders to Physicians from an Introspective Computer Medical Record: A Two-Year Randomized Trial. *Annals of Internal Medicine* 100, 130–138.
- [15] Bates, D.W., Cohen, M., Leape, L.L., Overhage, J.M., Shabot, M.M., and Sheridan, T. (2001) *Journal of the American Medical Informatics Association* 8, 299–308.
- [16] Oniki, T.A., Clemmer, T.P., and Pryor, T.A. (2003) The Effect of Computer-Generated Reminders on Charting Deficiencies in the ICU. *Journal of the American Medical Informatics Association* 10, 177–187.
- [17] Bates, D.W., Kuperman, G.J., Wang, S., Gandhi, T., Kittler, A., Volk, L., Spurr, C., Khorasani, R., Tanasijevic, M., and Middleton, B. (2003) Ten Commandments for Effective Clinical Decision Support: Making the Practice of Evidence-Based Medicine a Reality. *Journal of the American Medical Informatics Association* 10, 523–530.

Diversity in preoperative-Assessment Data Collection, a Literature Review

Leila AHMADIAN^{a,b,1}, Ronald CORNET^a, Wilton A van KLEI^c, Nicolette F. DE KEIZER^a

^a*Dept. of Medical Informatics, Academic Medical Center, University of Amsterdam Amsterdam, the Netherlands, b Kerman university of Medical Sciences, c dept of Anesthesiology, UMC, Utrecht*

Abstract: The appropriate anesthetic techniques and care during and after operation rely on data gathered during the preoperative assessment. Because various people are involved, standardization of this process is important. This paper provides a systematic literature review about which data items are collected in the preoperative assessment. Thirty-two relevant articles were found by PubMed search. To categorize data SNOMED CT concepts are used, resulting in 13 categories totaling 540 data items. The two largest categories of data were “past history of clinical finding”, and “physical examination procedure” with 251 and 75 data items respectively. Our study showed a high diversity of data items in the preoperative assessment. Because of the diversity of patients and treatment options available one undisputed preoperative assessment data set is hard to define. However, to solve the problem of exchangeability of the information at least anesthesiologists should use a same core set of data.

Key words: data set, standard, preoperative assessment, anesthesiology

Introduction

The rapidly changing health care environment requires appropriate, accurate and timely data to enhance communication between healthcare providers. It is important to collect data in a structured and standardized way to increase quality of data documentation. In the multidisciplinary setting of preoperative assessment standardization of the data is essential to facilitate effective communication between healthcare providers.

A proper preoperative assessment contributes to the reduction of postoperative incidents. Studies of perioperative morbidity and mortality repeatedly show that preoperative patients' conditions are significant predictors of intraoperative and postoperative morbidity and mortality[1,2]. The purposes of preoperative assessment are to estimate and reduce the mortality and morbidity risks associated with surgery, to determine required anaesthesia and equipment during the operation based on the patient's condition, to increase quality of care, to inform the patient about anaesthesia, and to obtain informed consent[3,4].

¹ Corresponding author: Department of medical informatics, Academic Medical Center, Room J1B-109, P.O.Box 22700 1100 DE Amsterdam, The Netherlands, Email: l.ahmadian@amc.uva.nl

In the past, the anaesthesiologist reviewed his or her own patients the night before surgery. Nowadays, this assessment is often done in preoperative clinics some days or weeks before surgery, in collaboration with other professionals such as specialized nurses. This new approach reduces the number of cancelled procedures. However, as the assessment is not necessarily performed by the anaesthesiologist who will actually administer anaesthesia, it reduces contact between care provider and patient. Consequently, this requires an increase in reliance on information that is obtained during the preoperative assessment: one should have detailed and ‘objective’ information right before the operation without performing part or all of a preoperative assessment again. Exchangeability of the information between different users and even different systems becomes increasingly important. Making use of existing standards such as terminological systems and reference information models contributes to this.

A preoperative assessment relies on collected information regarding a patient’s past history, physical examination, and clinical tests. Although the objective and the general domain of information in preoperative assessment are nearly the same in all settings, (the level of detail of) the information collected, however, remains unclear[3]. According to the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) standards[5], all surgical patients are required to have a history and physical examination documented in the patient record before surgery, but the components of data for those are not specified.

We performed a systematic literature review, with the main objective to identify data collected in the preoperative assessment and to investigate the degree of interdisciplinarity. This review functions as a first step towards determining which data items should be collected in the preoperative assessment. Its purpose is to provide a basis for designing a national core dataset for preoperative assessment in the Netherlands. This study is part of a larger effort to come to a standardized perioperative data set, which is performed in collaboration with the International Organization for Terminology in Anesthesia (IOTA).

1. Method

In order to find all articles in the last 10 years describing preoperative assessment datasets we used the PubMed database (January 1997 to June 2007). Only English language articles were included which fulfilled the search terms described in table 1. We used four sets of relevant key words and MeSH terms. Set 1 includes terms related to preoperative care, set 2 indicates assessment-related terms, and set 3 refers to possible ways or sources for data collection in preoperative assessment. Set 4 contains terms like “anaesthesia” or “preoperative” in the title of the article. We combined search terms using “OR” within the sets and “AND” among them. To check the recall of the search strategy we compared the search results with three relevant articles[5-7] and the references of a review article[3].

Two reviewers (LA and NdK) independently judged all titles and abstracts. Disagreements were discussed with a third reviewer (RC) and the final decision reflected consensus of all three reviewers. Articles were selected based on the following inclusion and exclusion criteria.

Editorials as well as original studies were included if they focused on collecting data in normal pre-anesthesia cases, or if they described preoperative data in one or more general categories of diseases, such as cardiac surgery. We excluded all articles

Table 1: Search terms used in the review. MeSH terms are in bold

Set 1	Set 2	Set 3	Set 4 (words in title)
Pre()operati*	evaluation	Medical History Taking	Anesthe*
Pre an(a)esthe*	preparation	Diagnostic Tests, Routine	Anaesthe*
Prean(a)esthe*	assessment	Physical Examination	Preopera*
Pre()surg*	Outcome Assessment	Questionnaires	Pre opera*
preoperative care	Risk Assessment	patient interview	Preanesthe*
“Outpatient Clinics, Hospital” AND (anesthe* OR anaesthe*)	Nursing Assessment	Patient(medical) history	Preanaesthe*

that described data collection for children, for patients with a specific disease or operation, and articles about patients at risk for specific complications. All articles focusing on predictors or risk factors of postoperative complications are excluded except those which explicitly mentioned that data about predictors or risk factors are commonly collected in preoperative assessment. Evaluation studies on necessity of specific tests, or specific parts of physical examination, and impact of different treatments or methods on risk factors, mortality, and morbidity were excluded. Studies were excluded that focused on organization of the preoperative clinic; time and cost management of operation room and the clinic; unless they described the data components. Systematic reviews, commentaries, and letters were excluded.

We extracted all data items part of the preoperative assessment described in the selected articles. As most of the articles had different ways for categorizing data we used SNOMED CT concepts to categorize the data.

To investigate the degree of interdisciplinarity, data about time between preoperative assessment and operation; disciplines involved in the preoperative process; and location for collecting data are extracted from each included article.

2. Results

In total we found 450 articles of which 63 articles were selected based on titles and abstracts. Based on the full-text review, 32 articles met our inclusion criteria. Preoperative assessment was performed between 30 days before surgery and the day of surgery, most of the studies stated one or more days before the surgery. Anesthesiologists, nurses, and other professionals such as surgeons, and consultants were involved in the preoperative assessment (figure 1). Three articles reported that the assessment forms should first be completed by the patient. In 47% of cases preoperative assessment was done in an outpatient clinic, 19% in hospitals and in 34% of cases the location was unknown.

We extracted 540 data items from the articles and used 13 categories to classify these data. Table 2 shows per main category, the number of articles mentioning this main category, or at least one of its data items, and the number of

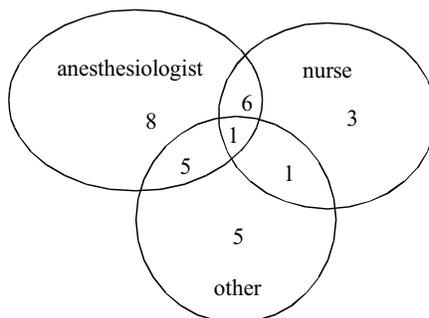


Figure 1: disciplines involved in preoperative assessment

Table 2: Frequency of collected data items per category extracted from literature

Category	Nr. of articles [references]	Nr. of data items in category	Data items, top 3 occurrence (n)
Demographic history detail	26[2,6,8-31]	11	Age (24) Sex (13) Weight/BMI (11)
Past history of clinical finding	31[2,5,6,8-35]	213	Diabetes (21) Cardiovascular diseases (21) Liver diseases (17)
Functional finding	16[2,5,6,9-12,18,19,24,25,27-30,35]	13	Exercise tolerance (9) Functional capacity (8) Prosthesis (5)
Behaviour finding	18[2,5,6,8-13,18,19,21,23,25,28-31]	4	Cigarette smoking and use of tobacco (16) Alcohol drinking (15) Illicit drugs (8)
Family history	13[2,5,6,9-11,14,18,25,28-30,35]	8	Anaesthesia- related problems (9) Malignant hyperthermia (4) Bleeding disorders (2)
Patient status observation	19[2,5,6,8,9,12,15-17,19,21,23-25,27-31]	27	Pregnancy (9) Informed consent (8) Comorbidities (7)
Review of medication	24[2,5,6,8-12,14-16,18-21,23-31]	68	Anticoagulant medications (15) Aspirin-containing medications (12) Diuretics (10)
Physical examination procedure	27[2,5,8-19,21,22,24-34]	75	Cardiovascular examinations (13) Blood pressure (11) Respiratory examinations (10)
Laboratory test	28[2,5,6,8-22,24,26,28-34,36]	53	Glucose (17) CBC (16) PT, PTT(15)
Diagnostic procedure	25[2,5,8-22,24,28-30,32-34,36]	11	ECG (21) Chest x ray (16) Echocardiography and pulmonary function tests (6)
Procedure	24[2,5,6,9,11-16,18-21,23-31,35]	11	Planned operation (13) Diagnoses (8) NPO status and type of anaesthesia (7)
Preoperative evaluation, anaesthesia	26[2,5,6,8-19,21-25,27-31,35]	31	Airway examinations(12) ASA class (11) Obesity, and anaesthesia complications history (10)
Administrative information	10[2,9,12,13,18,19,27-30]	15	Family or other support after surgery and during hospitalization (6) Information about how to contact with patient's family after surgery (4) Accompanying person, date of assessment, and transfer from (2)
Total	32	540	

data items. The right column shows the 3 most frequently used data items within a category.

3. Discussion and conclusion

This review showed a large diversity of data elements in preoperative assessment. Whereas each healthcare setting focuses on a limited set of data in preoperative assessment, we retrieved 540 data items in 13 categories. The categories “past history of clinical finding”, “physical examination procedure”, and “review of medication”

contained the largest number of data items. The observed interdisciplinarity (figure 1) and time delay (up to 30 days) between preoperative assessment and the operation emphasize the need for standardization of pre-operative data collection.

There are two limitations in our study. First, we may have missed some studies, because of the existing limitations in MeSH terms and lack of consideration of more relevant key words. We used the keywords of Van Klei's article[3] as a basis in our search strategy and expanded them to 4 sets of keywords (table 1). Given the extensive search strategy it is likely that most relevant articles were found. Furthermore inclusion of more papers will most likely merely strengthen the conclusion that there is a large diversity in preoperative assessment data collection. Second, studies that described general data for specific cases could have been missed, because we only addressed studies about normal pre-anesthesia cases, and cases regarding general categories of diseases.

The diversity of data collected by this review showed that almost each setting collected its own data set. Such diversity may also exist in other clinical settings. Designing a standard data set is necessary to overcome the dispersion of data among different settings. Although it is difficult to introduce an undisputed data set for the preoperative assessment, interoperability requires a standard core data set. It would facilitate patient referrals across health care settings, and help healthcare providers to have a same understanding of patients' conditions. This would increase the quality of care. Based on their clinical needs institutions could add more data to the core data set. To facilitate the exchangeability of data and effective communication among health care providers, we recommend using standard terminological systems and information models such as SNOMED CT and LOINC to give value to the data items.

This study demonstrated the diversity in collected pre-operative assessment data and revealed the most frequently used data elements and the core categories of preoperative data. The results of this study will be used by experts in order to design a Dutch national data set.

Reference List

- [1] M. F. Roizen, Preoperative evaluation of patients: a review, *Ann. Acad. Med. Singapore*, 23 (1994) 49-55.
- [2] A. Bray, Preoperative nursing assessment of the surgical patient, *Nurs. Clin. North Am.*, 41 (2006) 135-50, v.
- [3] W. A. van Klei, D. E. Grobbee, C. L. Rutten, P. J. Hennis, J. T. Knappe, C. J. Kalkman, and K. G. Moons, Role of history and physical examination in preoperative evaluation, *Eur. J. Anaesthesiol.*, 20 (2003) 612-618.
- [4] M. F. Roizen, More preoperative assessment by physicians and less by laboratory tests, *N. Engl. J. Med.*, 342 (2000) 204-205.
- [5] F. A. Michota and S. D. Frost, The preoperative evaluation: use the history and physical rather than routine testing, *Cleve. Clin. J. Med.*, 71 (2004) 63-70.
- [6] N. H. Badner, R. A. Craen, T. L. Paul, and J. A. Doyle, Anaesthesia preadmission assessment: a new approach through use of a screening questionnaire, *Can. J. Anaesth.*, 45 (1998) 87-92.
- [7] E. Lew, D. J. Pavlin, and L. Amundsen, Outpatient preanaesthesia evaluation clinics, *Singapore Med. J.*, 45 (2004) 509-516.
- [8] T. Ho and A. E. Brissett, Preoperative assessment of the aging patient, *Facial. Plast. Surg.*, 22 (2006) 85-90.
- [9] B. J. Sweitzer, Preoperative medical testing and preparation for ophthalmic surgery, *Ophthalmol. Clin. North Am.*, 19 (2006) 163-177.
- [10] F. A. Michota, Jr., The preoperative evaluation and use of laboratory testing, *Cleve. Clin. J. Med.*, 73 Suppl 1 (2006) S4-S7.
- [11] A. K. Ghosh, Preoperative evaluation for non-cardiac surgery, *J. Assoc. Physicians India*, 53 (2005) 306-311.

- [12] AORN Guidance Statement: preoperative patient care in the ambulatory surgery setting, *AORN J.*, 81 (2005) 871-878.
- [13] A. Ferrando, C. Ivaldi, A. Buttiglieri, E. Pagano, C. Bonetto, R. Arione, L. Scaglione, E. Gelormino, F. Merletti, and G. Ciccone, Guidelines for preoperative assessment: impact on clinical practice and costs, *Int. J. Qual. Health Care*, 17 (2005) 323-329.
- [14] T. M. Halaszynski, R. Juda, and D. G. Silverman, Optimizing postoperative outcomes with efficient preoperative assessment and management, *Crit Care Med.*, 32 (2004) S76-S86.
- [15] A. E. Greer and M. G. Irwin, Implementation and evaluation of guidelines for preoperative testing in a tertiary hospital, *Anaesth. Intensive Care*, 30 (2002) 326-330.
- [16] S. P. Fischer, Cost-effective preoperative evaluation and testing, *Chest*, 115 (1999) 96S-100S.
- [17] R. H. Haug and R. L. Reifeis, A prospective evaluation of the value of preoperative laboratory testing for office anesthesia and sedation, *J. Oral Maxillofac. Surg.*, 57 (1999) 16-20.
- [18] N. Girard, Preoperative assessment, *Nurse Pract. Forum*, 8 (1997) 140-146.
- [19] S. T. McGillis and U. Stanton-Hicks, The preoperative patient evaluation: preparing for surgery, *Dermatol. Clin.*, 16 (1998) 1-15.
- [20] E. Capdenat Saint-Martin, P. Michel, J. M. Raymond, H. Iskandar, C. Chevalier, M. N. Petitpierre, L. Daubech, M. Amouretti, and P. Maurette, Description of local adaptation of national guidelines and of active feedback for rationalising preoperative screening in patients at low risk from anaesthetics in a French university hospital, *Qual. Health Care*, 7 (1998) 5-11.
- [21] M. A. Starsnic, D. M. Guarnieri, and M. C. Norris, Efficacy and financial benefit of an anesthesiologist-directed university preadmission evaluation center, *J. Clin. Anesth.*, 9 (1997) 299-305.
- [22] M. Hunt, Nurses can enhance the pre-operative assessment process, *Nurs. N. Z.*, 12 (2006) 20-21.
- [23] S. Charuluxananan, S. Suraseranivongse, Y. Punjasawadwong, W. Somboonviboon, T. Nipitsukarn, T. Sothikarnmanee, M. Vasinanukom, T. Werawatganon, S. Tanudsintum, V. Lekprasert, and T. Hintong, The Thai Anesthesia Incidents Study (THAI Study) of anesthetic outcomes: I. Description of methods and populations, *J. Med. Assoc. Thai.*, 88 Suppl 7 (2005) S1-13.
- [24] W. A. van Klei, P. J. Hennis, J. Moen, C. J. Kalkman, and K. G. Moons, The accuracy of trained nurses in pre-operative health assessment: results of the OPEN study, *Anaesthesia*, 59 (2004) 971-978.
- [25] G. R. de Oliveira Filho and L. Schonhorst, The development and application of an instrument for assessing resident competence during preanesthesia consultation, *Anesth. Analg.*, 99 (2004) 62-69.
- [26] B. M. Alsumait, S. A. Alhumood, T. Ivanova, M. Mores, and M. Edeia, A prospective evaluation of preoperative screening laboratory tests in general surgery patients, *Med. Princ. Pract.*, 11 (2002) 42-45.
- [27] S. Ausset, H. Bouaziz, M. Brosseau, B. Kinirons, and D. Benhamou, Improvement of information gained from the pre-anaesthetic visit through a quality-assurance programme, *Br. J. Anaesth.*, 88 (2002) 280-283.
- [28] E. Clark, Preoperative assessment. Primary care work-up to identify surgical risks, *Geriatrics*, 56 (2001) 36-40.
- [29] L. B. Latham, Preanesthetic evaluation, *Dent. Clin. North Am.*, 43 (1999) 217-229.
- [30] D. Dunn, Preoperative assessment criteria and patient teaching for ambulatory surgery patients, *J. Perianesth. Nurs.*, 13 (1998) 274-291.
- [31] D. J. Essin, R. Dishakjian, V. L. deCiutiis, C. D. Essin, and S. N. Steen, Development and assessment of a computer-based preanesthetic patient evaluation system for obstetrical anesthesia, *J. Clin. Monit. Comput.*, 14 (1998) 95-100.
- [32] F. Y. Ajimura, A. S. Maia, A. Hachiya, A. S. Watanabe, M. P. Nunes, M. A. Martins, and F. S. Machado, Preoperative laboratory evaluation of patients aged over 40 years undergoing elective non-cardiac surgery, *Sao Paulo Med. J.*, 123 (2005) 50-53.
- [33] S. Mantha, M. F. Roizen, J. Madduri, Y. Rajender, N. K. Shanti, and K. Gayatri, Usefulness of routine preoperative testing: a prospective single-observer study, *J. Clin. Anesth.*, 17 (2005) 51-57.
- [34] H. Kinley, C. Czoski-Murray, S. George, C. McCabe, J. Primrose, C. Reilly, R. Wood, P. Nicolson, C. Healy, S. Read, J. Norman, E. Janke, H. Alhameed, N. Fernandes, and E. Thomas, Effectiveness of appropriately trained nurses in preoperative assessment: randomised controlled equivalence/non-inferiority trial, *BMJ*, 325 (2002) 1323.
- [35] W. G. Hilditch, A. J. Asbury, E. Jack, and S. McGrane, Validation of a pre-anaesthetic screening questionnaire, *Anaesthesia*, 58 (2003) 874-877.
- [36] B. A. Finegan, S. Rashid, F. A. McAlister, and P. O'Connor, Selective ordering of preoperative investigations by anesthesiologists reduces the number and cost of tests, *Can. J. Anaesth.*, 52 (2005) 575-580.

Representing Health, Disorder and their Transitions by Digraphs

Carl-Fredrik BASSØE, PhD (medicine), PhD (informatics);

Norwegian Centre for Electronic Medical Records, Institute for Neuroscience, Faculty of Medicine, University of Science and Technology, Trondheim, Norway and the

PROMED Institute, Bergen, Norway

Abstract. In this study clinical decision making (CDM) is formalized by representing the aetiology and the human body by one directed graph (digraph) and using standard digraph operators (change, add, delete, contract) to model transitions between health and disorder. All nodes of the digraph have the same composite structure $\langle localization, carrier, agent \rangle$. For example, an aetiology node is $\langle Ganges\ river, water, Vibrio\ cholera \rangle$. Paths in the aetiology subdigraph model epidemiological spread. Virulent paths model the entrance into and damage of aetiological agents on the body. Pathogenetic mechanisms make out internal pathways between organs and between cells, and is sharply discriminated from the aetiology. CDM is based on recognizing the difference between healthy and disordered digraphs. The result is a novel and powerful approach to CDM.

Keywords. health, disease, digraph, digraph operators, clinical decision-making

Introduction

A successful formal theory for clinical decision-making (CDM) may aid practical clinical work and facilitate the development of medical expert systems, but formalizing CDM is a hard unsolved problem.

A unified foundation for CDM needs explicit discrimination between health and disease based on standard medical literature and a formal socio-psycho-biological model. Diagnoses can be constructed by the equation $d:=e+o+p$, where d , e , o and p hold names of a diagnoses, an aetiology, an organ and a pathogenetic mechanism, respectively.^{1,2} This work concerns a mapping of the latter three entities to a directed graph (digraph).

1. Background

Graphs and networks are commonly used to describe, manipulate and illustrate healthy and clinical structures and functions.^{3,4,5,6} Graph theory applies to anatomy⁷, pedigrees⁸, epidemiology⁹ and metabolic pathways¹⁰. DNA, RNA and proteins can be represented

by grammars^{11,12} and generated on directed graphs¹³. Mammalian cells are modelled as nodes connected by regulators, metabolites and ions¹⁴. Furthermore, neural networks and pathways are analyzed using graph theory¹⁵. Central aspects of language are represented by digraphs^{16,17} that underlie data traffic in social networks¹⁸. Graphs are basic to medical ontologies that are used in clinical guidelines and aids decision support¹⁹. Even though blocks in cell differentiation may arrest apoptosis and lead to leukaemia²⁰ and changes in cell differentiation are hallmarks of myelodysplasia²¹ the altered structure of affected body parts are rarely taken into account in medical expert systems (MES). MES can be designed on graphs²² or simply use sets of symptoms and signs associated with diseases and syndromes²³. Semantic networks and graphs are also used in MES^{3,15,20,24,25,26}, but such MES are constrained to narrow clinical domains. CDM can be based on separate analysis of the aetiology, the affected body part and the pathogenesis²⁷, but formal tools to model the transitions between health and disorder are unavailable.

2. Methods

The standard notation and definitions used here are from Gross²⁸. A directed graph (digraph) X consists of a set of nodes $N(X)$ and a set of arches $A(X)$. Each arch is associated with one or two nodes. Two nodes are incident if they are joined by one edge. Two incident arches share a node.

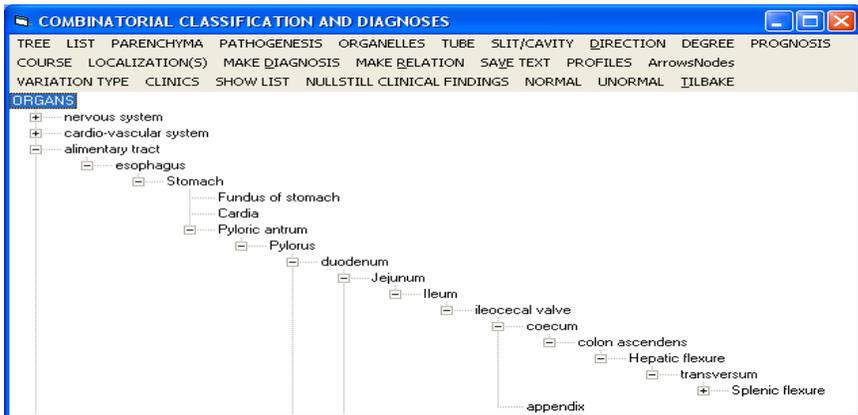


Figure 1. Implementation of systematic anatomy of the oesophago-gastro-intestinal tract.

Arches $f: A \rightarrow B$ point from a head node A to a tail node B . The arch f may represent a structural or functional relationship between A and B . Pathways are organized by sequences of nodes and arches. In figure 1 nodes are implemented by records with names, and pointers implement the arches between the records. CDM involving the origin of a rectal bleeding, for example, proceeds by following the path from the rectum in proximal direction and examining each location for a bleeding source – in accordance with common clinical practice.

Digraph operators are used to model transitions between health and disorder. The symbols \downarrow , $-$ and $+$ mean contraction, delete and add, respectively (see fig. 2). Also, $f \mapsto g$ means that f changes into g . The set of operators can be extended².

3. Results

Let H and D represent a healthy and a disordered digraph, respectively. The effect of the $+$ operator is illustrated in figure 2. Nodes represent cardiovascular tubes and heart cavities. Arches represent blood flow through tubes, cavities and heart valves. Adding the arch asd to H (Fig. 2H) in direction from the left to the right atrium gives rise to a disordered digraph $D=H+asd$ representing atrium septum defect (Fig. 2D). The removal of RV by contraction, formally given by $H \downarrow RV$, leads to union of RA and LA , i.e. a heart with one atrium and one ventricle. Deletion of one arch t, p, m or a in figure 2 is incompatible with life. Clearly, the operators model disease processes.

Here, CDM is based on recognizing differences between H and D . For example, inborn errors of metabolism are recognized by the presence of an enzyme node in a metabolic pathway in H and its absence or malfunction in D .

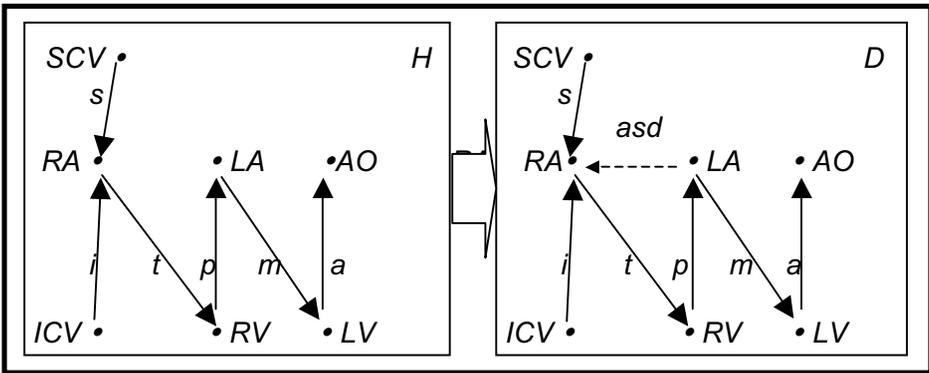


Figure 2. Digraph representing circulation through a healthy heart (H) and atrium septum defect (D). SCV : superior caval vein, ICV : inferior caval vein, RA : right atrium, LA : left atrium, RV : right ventricle, LV : left ventricle, AO : aorta. Arrows show flow through ICV (i), SCV (s) the tricuspid valve (t), pulmonary valve (p), mitral valve (m), aortic valve (a) and asd .

3.1. Graph operators create disordered digraphs

Formally, digraph operators transform H into D . The deletion of a node n from H gives rise to the digraph $D=H-n$ with node set $N(D)=N(H)-n$. Let all arches pointing from or to n be in the set f_n . Then $D=H-n$ implies $A(D)=A(H)-f_n$, $N(H)-N(D)$ and $A(H)-A(D)$ are the sets of nodes and arches, respectively, that are in H , but not in D . The differences discriminate H from D and are sufficient for CDM regarding deleted nodes. Examples vary from traumatic loss of body parts to agenesis (anencephaly) and aplasia.

Deleting an arch f from H gives $A(D)=A(H)-f$ while $N(D)=N(H)$. Let $f:A \rightarrow B$ model blood flow from artery A to artery B . An trauma between A and B that eliminates blood flow is modelled by the deletion of f . Let A and B be immature cells in a cell line and the arrow between them denote cell differentiation. Deleting f blocks apoptosis and arrests differentiation. With leukocytes the deletion results in acute leukaemia.

The change (C) an arch f to g ($f \rightarrow g$) gives rise to numerous disorders:

1. Let $(D,g)=C(H,f)$ imply $f \in A(H)$, $f \notin A(D)$, $g \notin A(H)$ and $g \in A(D)$. Typical examples are disorders of the concentration f_S of substance S in peripheral blood. Thus, $g_S > f_S$ implies ‘hyperSemia’ and $g_S < f_S$ means ‘hypoSemia’ where S is the substance

name. If $S = \text{'glyc'}$ then 'hyperSemia' means 'hyperglycaemia'. If $f \rightarrow g$ reflects altered function of the node F , then $g \prec f$ leads to 'hypo F function'. If $f = \text{'tension'}$ and $F = \text{'arterial'}$ then $g \succ f$ means 'arterial hypertension'.

2. Let l in $(D, g) = l(H, f)$ invert the direction of f . If f represents food transport through the oesophagus then g denotes reversed transport, i.e. vomiting.

3. Let R in $(D, g) = R(H, f)$ replace the arch f in H with the undirected edge e . If $f: A \rightarrow B$ represents transport from A to B then $e: A \rightarrow B$ implies $e(A) = 0$. Occlusions and atresia are typical examples.

4. Let A and B be immature cells in a cell line in H . Changes in cell differentiation ($f \rightarrow g$) given by $f: A \rightarrow B$ to $g: A \rightarrow Y$ may model dysplasia, e.g. myelodysplasia.

Change a node. Let X be a node in H . Let a change $X \rightarrow Y$ be induced by $(D, Y) = F(H, X)$. For example, F_A decreases the size of an organ or organ part (atrophy) and F_Y increases their size (hypertrophy).

Insertion of a node $Y \notin N(H)$ into H gives rise to the digraph D with node set $N(D) = N(H) + Y$. The insertion requires the addition of at least one arch $g_D \notin A(H)$. Since g_D connects to a node in H and the node Y in D we also need $A(D) = A(H) + g_D$. Typical examples are tumours and abscesses.

Insertion of an arch $g \notin A(H)$ gives rise to the digraph D with edge set $A(D) = A(H) + g$ and node set $N(D) = N(H)$. The insertion of g from node A to node B alters the adjacency of the two nodes. An examples is the accessory electrical pathway in Wolff-Parkinson-White syndrome where g models the pathological extra electrical circuit. In bleeding aneurysms g models the blood stream through the arterial wall.

The operators contract, delete and add lead to $|N(H)| \neq |N(D)|$, $|A(H)| \neq |A(D)|$ or both, and CDM involves the identification of nodes and arches that discriminate D from H . All remaining D 's have at least one changed node or arch. Now CDM depends on pair wise comparisons of nodes and arches in D and H and recognizing the difference(s).

3.2. Aetiology epidemiology and virulence

The aetiology is the set of environmental agents that cause disease. The agents arise externally to organisms in time (heredity) or space. The nodes of the aetiology digraph are triplets $\langle \text{localization, carrier, agent} \rangle$. Hereditary transmission of disease is described using pedigrees with arches like $M: \langle \text{Father, cell nucleus, mutated DNA} \rangle \rightarrow \langle \text{Son, cell nucleus, mutated DNA} \rangle$. Inborn errors of metabolism are represented by DNA nucleotide contractions, deletions, insertions and changes (point mutations). Epidemiological spread is modelled by arches such as $M: \langle \text{Asia, chicken, influenza virus} \rangle \rightarrow \langle \text{Europe, chicken, influenza virus} \rangle$, which extends to chemicals, allergens, radiation and harmful mechanical agents. Person are in nodes in the social digraph. Its arches model messages. The direction of arches is from motor outputs (behaviour) to sensory inputs. Certain message patterns cause mental disorder.

Translocation of agents into the organism is represented by a set of arches $T: \mathbf{A} \rightarrow \mathbf{O}_p$ where \mathbf{A} and \mathbf{O}_p refer to the agent and the primary affected organ, respectively. A typical example is translocation of hepatitis A virus (HAV) from a water source to the liver given by $T: \langle \text{Ganges, water, HAV} \rangle \rightarrow \langle \text{Liver, hepatocyte, HAV} \rangle$. The damage caused by HAV is given by $G: \langle \text{Liver, normal hepatocytes, HAV} \rangle \rightarrow \langle \text{Liver, damaged hepatocytes, HAV} \rangle$. Virulence denotes both the propensity to translocate and the degree of damage and is represented by $V = GT$ meaning T followed by G .

3.3. Pathogenesis

Body nodes represent organs, organ parts, tissues and cells. By definition \mathbf{O}_P are hit by aetiological agents². Damaged \mathbf{O}_P modify other body nodes through pathogenetic mechanisms (P_o and P_i) that involve the mesenchyme, immunological reactions, intercellular metabolites and regulators, and the vascular and neural systems. Thus, the pathogenetic relation between the \mathbf{O}_P and secondary affected nodes (\mathbf{O}_S) is given by $P_o: \mathbf{O}_P \rightarrow \mathbf{O}_S$ and $P_i: \mathbf{O}_S \rightarrow \mathbf{O}_P$. For example, if P_o represents signals to the bone marrow and P_i the path of neutrophils from \mathbf{O}_S to \mathbf{O}_P then $P_i P_o$ represents accumulation of neutrophils in \mathbf{O}_P . In effect, P_o and P_i represent disorders ranging from immune reactions and metabolic disturbances to psychosomatic disorders².

4. Discussion and conclusion

This study shows that transitions from health to disease can be represented by digraph operators. The differences between healthy and disordered digraphs allows their discrimination. This provides new formal basis for CDM.

CDM needs a mechanism to trace the origin of processes such as pareses and ischemia. This problem is solved by following systematic anatomical paths (Fig. 1). Any representation useful to CDM must model the causes of clinical disorders, and describe disordered structure of internal objects and relations. The digraph fulfils these needs by modelling the aetiology, organ disorders and pathogenetic relations. Minds connect to social networks. Therefore, the representation extends to psychosomatic disorders. Endocrine disturbances are traced in metabolic paths² of the pathogenesis.

Causes of stenoses are usually independent of their localization. Assume that there are c causes and a localizations. Then conventional CDM involves probing a search space of $a \cdot c$ elements. In contrast, partition CDM into the independent analysis of arches and processes. This reduces the number of elements to be investigated to $a+c$ and simplifies CDM². The argument extends to partitioning the search space given by all diagnoses into separate lists of elements from the aetiology, disordered organs and pathogenetic mechanisms. Thereafter diagnoses are constructed by $d:=e+o+p$. This is an algorithmic approach that profoundly simplifies CDM.

The rule-based systems and semantic networks referred to above come close to the present work. However, the former are informal approaches to CDM whereas the present is formal. The digraph operators used here are well defined, limited in number, and cover changes in nodes and arches.

Graph theory has important limitations. The dynamics of (patho)physiology and shortcuts in clinical reasoning are difficult to model by digraphs alone. Formal category theory seems a promising extension in this domain and promising tests have been performed².

In conclusion, I have shown that digraphs represent functions and structures underlying CDM. Digraph operators represent the transitions between health and disorder. This is a novel and powerful approach to CDM.

References

- [1] C-F.Bassøe, A data structure for decision support systems, medical expert systems and clinical decision making, *MEDINFO 2007*, HISA Ltd Australia (2007), P350.
- [2] C-F.Bassøe, Combinatorial clinical decision-making, PhD thesis, *Institute for Information Science and Media, Faculty of Social Sciences*, University of Bergen, Norway, 2007.
- [3] G.Bigras, R.Marcelpoil, E.Brambilla, G.Brugal, Cellular sociology applied to neuroendocrine tumors of the lung: quantitative model of neoplastic architecture, *Cytometry* **24** (1996), 74-74.
- [4] R.Milo, S.Shen-Orr, S.Itzkovitz, N.Kashtan, D.Chklovskii, U.Alon, Network motifs: simple building blocks of complex networks, *Science* **298** (2002), 824-827.
- [5] Z.N.Oltvai, A-L.Barabási, Life's complexity pyramid, *Science* **298** (2002), 763-763.
- [6] A.H.Y.Tong, B.Drees, G.Nardelli, G.D.Bader, B.Brannetti, L.Castagnoli, et al, A combined experimental and computational strategy to define protein interaction networks for peptide recognition molecules, *Science* **295** (2002), 321-321.
- [7] T.F.Hayamizu, M.Mangan, J.P.Corradi, J.A.Kadin, M.Ringwald, The Adult Mouse Anatomical Dictionary: a tool for annotating and integrating data, *Genome Biol* **6** (2005), R29.
- [8] A.Piccolboni, D.Gusfield, On the complexity of fundamental computational problems in pedigree analysis, *J Comput Biol* **10** (2003), 763-773.
- [9] I.M. Longini, M.E.Halloran, Strategy for distribution of influenza vaccine to high-risk groups and children, *Am J Epidemiol* **161** (2005), 303-303.
- [10] X.Q.Xia, M.J.Wise, DiMSim A discrete-event simulator of metabolic networks, *J Chem Inf Comput Sci* **43** (2003), 1011.
- [11] M.A.Jiménez-Montaña, Formal languages and theoretical molecular biology, In, B.Goodwin, P.Daunders P (Eds.), *Theoretical biology*, Johns Hopkins University Press, Baltimore (1992), 199.
- [12] J.Collado-Vides, *Towards a grammatical paradigm for the study of the regulation of gene expression*. In, B.Goodwin B, Daunders P (Eds.), *Theoretical biology*, Johns Hopkins University Press, Baltimore (1992), 211-24.
- [13] R.F.C.Walters, *Categories and computer science*, Cambridge University Press, 1991.
- [14] A.Ma'ayan, S.L.Jenkins, S.Neves, A.Hasseldine, E.Grace, B.Dubin-Thaler, N.J.Eungdamrong, G.Weng, P.T.Ram, J.J.Rice, A.Kershenbaum, G.A.Stolovitzky, R.D.Blitzer, R.Iyengar, Formation of regulatory patterns during signal propagation in a Mammalian cellular network, *Science* **309** (2005), 1078-1083.
- [15] J.F.Gomez, F.J.Lopera, A topological hypothesis for the functional connections of the cortex A principle of the cortical graphs' based on the neuroimaging, *Med Hypotheses* **53** (1999), 263-262.
- [16] D.A.Cohen, Introduction to computer theory, *John Wiley*, New York, 1986.
- [17] D.Jurafsky, J.H.Martin, Speech and language processing, *Prentice Hall*, Englewood, NJ, 2000.
- [18] M.E.J.Newman, The structure and function of complex networks, *SIAM Review* **45** (2003), 167-256.
- [19] S.L.Achour, M.Dojat, C.Rieux, P.Bierling, E.Lepage, A UMLS-based knowledge acquisition tool for rule-based clinical decision support system development., *J Am Med Inform Assoc* **8** (2001), 351-360.
- [20] D.V.Catenacci, G.J.Schiller, Myelodysplastic syndromes: a comprehensive review., *Blood Rev* **19** (2005), 301-319.
- [21] A.C.Fisher, A.Chandna, I.P.Cunningham, The differential diagnosis of vertical strabismus from prism cover test data using an artificially intelligent expert system., *Med Biol Eng Comput* **45** (2007), 689-693.
- [22] C.-F.Bassøe, O.Bruserud, I.F.Pryme, A.Vedeler, Ribosomal proteins sustain morphology, function and phenotype in acute myeloid leukemia blasts, *Leuk Res* **22** (1998), 329-329.
- [23] H.Yan, Y.Jiang, J.Zheng, B.Fu, S.Xiao, C.Peng, The internet-based knowledge acquisition and management method to construct large-scale distributed medical expert systems., *Comput Methods Programs Biomed* **74** (2004), 1-10.
- [24] J.H.Sowa, Semantic networks, <http://www.jfsowa.com/pubs/semnet.htm>, 2006, accessed 02/28/2008.
- [25] A.Pommert, K.H.Höhne, E.Burmester, S.Gehrmann, R.Leuwer, A.Petersik, B.Pflesser, U.Tiede, Computer-based anatomy a prerequisite for computer-assisted radiology and surgery., *Acad Radiol* **13** (2006), 104-112.
- [26] S.Liu, J.L.Gaudiot, V.Cristini, Prototyping virtual cancer therapist (VCT): a software engineering approach., *Conf Proc IEEE Eng Med Biol Soc* **1** (2006), 5424-5427.
- [27] C.-F.Bassøe, Automated diagnoses from clinical narratives: A medical system based on computerized medical records, natural language processing and neural network technology, *Neural Networks* **8** (1995), 313-313.
- [28] J.Gross, J.Yellen, Graph theory and its applications, *CRC Press*, Inc, Boca Raton, Florida, 1999.

Characterizing the Dimensions of Clinical Practice Guideline Evolution

Jacques BOUAUD^a, Brigitte SÉROUSI^b

^a*AP-HP, DSI, STIM, Paris, France; INSERM, UMR_S 872, eq. 20, Paris, France*

^b*Université Paris 6, UFR de Médecine, Paris, France; AP-HP, Hôpital Tenon, Département de Santé Publique, Paris, France; APREC, Paris, France*

Abstract. The ever growing pace at which medical knowledge is produced requires clinical practice guidelines (CPGs) to be regularly updated. Since clinical decision support systems (CDSSs) are effective means to implement guidelines in routine care, they have to be revised as their knowledge sources evolve. From one version to another, some parts are kept unchanged whereas others are more or less modified. We propose to characterize formally the different dimensions of recommendation evolution in two successive guideline versions from the knowledge modelling perspective. Each atomic recommendation is represented as a rule connecting a clinical condition to recommended action plans. Using subsumption-based comparisons, seven evolution patterns were identified: *No change*, *Action plan refinement*, *New action plan*, *Condition refinement*, *Recommendation refinement*, *New practice*, and *Unmatched recommendation*. The method has been evaluated on French bladder cancer guidelines in the revisions of 2002 and 2004.

Keywords. Clinical practice guideline updating, clinical decision support system, knowledge base revision, bladder cancer management

Introduction

Significant practice variations are frequently observed and numerous medical errors occur in most medical institutions. Thus, there has been a worldwide interest for clinical practice guidelines (CPGs) expected to improve quality of care by disseminating evidence-based practice and assist practitioners in the state-of-the-art management of patients. However, CPGs are usually produced as textual documents, the dissemination of which is poorly effective in changing the behavior of physicians.

On the contrary, clinical decision support systems (CDSSs) seem to be efficient to influence healthcare professionals in the adoption of CPGs [1]. Defined as any software in which characteristics of individual patients are matched to a computerized knowledge base (KB) for the purpose of generating patient-specific assessments, or recommendations, CDSSs rely on computerized versions of original textual CPGs. However, despite the development of numerous guideline representation formalisms, the translation of textual CPGs into computerized KBs is still difficult and expensive [2]. In addition, KBs are often difficult to compare to original documents.

On the other side, medical knowledge is continuously increasing and the very meaning of “state of the art” or “evidence” are relative and evolving notions. The consequence is that CPGs are necessarily involved in a life-cycle and must be updated regularly by health professional societies or national health agencies in charge of their development. In France, the “Haute Autorité de Santé”, (www.has-sante.fr), is responsible of CPG development, dissemination and updating. However, as opposed to living guidelines [3, 4], updating CPGs in France seldom capitalizes on the previous version of the same document, and revised versions of CPGs are often totally new CPGs. As a consequence, when updating guideline-based CDSSs to account for guideline revision, KBs have to be re-developed from the new CPG version.

The aim of this paper is to characterize the differences existing between two successive versions of computerized KBs and propose a formal characterization of knowledge evolution underlying CPG updating. The method has been applied to bladder cancer guidelines developed by the French Association for Urology (“Association Française d’Urologie”) in the revisions of 2002 and 2004.

1. Background

1.1. Document vs. knowledge-base centric approaches to CPG modelling

Most guideline tools focus on CPG documents as the starting point for guideline annotation and KB creation. Document-centric approaches such as the Guideline Element Model (GEM [5]), the Digital electronic Guideline Library (DeGeL [6]), Stepper [7], and the Guideline Markup Tool (GMT [8]) use guideline documents as the knowledge source and operate a gradual process of creating KBs from documents. Most of the time, those markup-based tools use XML-based documents to link textual descriptions with structured guideline models. The relationship between the semi-formal model of marked text parts and the original document is thus preserved. In model-centric approaches, the underlying conceptual model of guidelines is build by domain experts on the basis of their interpretation to clear text ambiguities and complete document incompleteness [2]. In both approaches, the final step is the translation of the formal guideline model into a guideline representation formalism (GLIF, Asbru, PROforma, etc.) and the building of a KB.

On the other hand, SAGE aims at promoting a knowledge-base-centric approach. In order to provide a human-comprehensible view of the guideline knowledge encoded in Protégé ontologies and KBs, the SAGE project developed a method for translating portions of a guideline KB into a legible format. Thus, SAGE uses KBs as the main guideline representation format and generates guideline documents from its models. Eriksson *et al.* [9] proposed to automate the document production from a computer interpretable guideline KB while combining documents with semantic information from the guideline KB. Although they used the SAGE guideline model, the approach is applicable for KBs that have well-defined semantics.

1.2. CPG updating

In document-centric approaches, when a guideline document is updated, updating the corresponding KB is well-handled in the case of “living guidelines” [3, 4]. Indeed,

living guidelines are incrementally built from previous versions and highlight the information that was changed: modifications of every revision are annotated, time-stamped and marked by arrows. In this case, differences between two successive versions of the document are easily recognizable. Unchanged parts of the text correspond to already modelled parts in the KB and are thus inherited. Only new textual parts have to be modelled.

In the case of “non-living guidelines”, even using natural language processing, identifying what was changed in the new textual version is not easy and the new guideline-based KB has often to be built from scratch. In addition, the development of a new version of non-living guidelines usually occurs every 4 to 5 years, or according to the availability of new evidence impacting clinical practice. This process may take up to 2 years, so that CPGs can be out of date as they are published. Thus, KBs may be more frequently updated on an ongoing basis as new evidence becomes available.

We propose a method to compare two KBs built from two successive revisions of the same CPGs. Beyond highlighting the differences between the two KBs, the method provides a formal characterization of practice guideline knowledge evolution.

2. Method

2.1. Notations and identity

CPGs can be considered as a set of atomic recommendations, noted $\{R_i\}_i$. Each recommendation R_i is characterized by a pair (S_i, P_i) such as $S_i \Rightarrow P_i$, denoting that an action plan P_i is recommended in the clinical situation S_i . A clinical situation S_i is defined by a set of instantiated decision variables: $S_i = \{c_{i,n}\}_n$. An action plan P_i is defined by an ordered sequence of medical actions: $P_i = (a_{i,m})_i$.

The basic comparison between 2 recommendations is based on identity: R_i and R_j are identical if and only if their clinical situations are identical (same sets of decision variables) and their action plans are identical (same sets of actions in the same order):

$$R_i = R_j \Leftrightarrow (S_i = S_j) \wedge (P_i = P_j)$$

2.2. Similarity and subsumption

Even when not identical, recommendations may share some commonality. We propose to use subsumption functions to compare recommendations at a more abstract level.

In “ontological” abstraction, abstract concepts have to be identified and a mapping, noted *abst*, between basic elements and their abstract class is built. To avoid the problem of over-generalisation and, for practical reasons, we limited the abstraction by specifying for each concept only one abstract concept relevant for the application. Abstraction is performed on decision variables and on actions. For instance, $abst(gemcitabin) = abst(cysplatin) = chemotherapy$ so that gemcitabin and cysplatin are both considered as chemotherapies, but are not comparable to any radiotherapy or surgery. Action plan similarity, noted *sim*, is a boolean function taking 2 action plans for arguments. We consider that 2 action plans are similar if they involve the same actions. Since action plans are sequences, it consists in not considering order, and similarity is therefore defined as set identity: $sim(P_i, P_j) \Leftrightarrow (P_i \subseteq P_j) \wedge (P_j \subseteq P_i)$.

Clinical situations can be compared using a structure-based subsumption function, noted *subsum*, taking S_i and S_j as arguments, and returning true when S_j is more specific

than S_i , typically by involving additional decision variables. A simple subsumption function is then to consider set inclusion: $subsum(S_i, S_j) \Leftrightarrow (S_i \subseteq S_j)$.

Similarity and subsumption functions can be enhanced using abstraction on their arguments. We introduced the two functions Sim and $Subsum$ defined as follows:

$$\begin{aligned} Sim(P_i, P_j) &= sim(abst(P_i), abst(P_j)) \\ Subsum(S_i, S_j) &= subsum(abst(S_i), abst(S_j)) \end{aligned}$$

2.3. Evolution patterns

Using the notations previously introduced and the comparison algorithm described in figure 1 between 2 recommendations, assuming that R_j is an evolution of R_i , we formally obtain seven evolution patterns that are qualitatively described as follows:

1. *No change*: recommendation R_i remains unchanged; R_i and R_j are identical
2. *Action plan refinement*: both clinical situations are identical, but their action plans, though different, share common properties according to the plan similarity function, e.g. $S_i^{2002} = S_j^{2004} = \{no\ complete\ tumor\ resection, invasive\ tumor, non\ metastatic\ disease, N2, no\ contraindication\ to\ chemotherapy\}$, $P_i^{2002} = \{chemotherapy\}$ and $P_j^{2004} = \{adjuvant\ chemotherapy\}$.
3. *New action plan*: both clinical situations are identical, but their action plans are not considered similar. R_j recommends a new action plan in the same situation e.g. $S_i^{2002} = S_j^{2004} = \{no\ complete\ tumor\ resection, invasive\ tumor, non\ metastatic\ disease, N2, T4b, no\ contraindication\ to\ chemotherapy\}$, $P_i^{2002} = \{radiotherapy\}$ and $P_j^{2004} = \{follow-up\}$.
4. *Condition refinement*: clinical situation S_j is more specific than S_i , i.e. R_j is more restrictive than R_i , but the same action plan is recommended in both situations e.g. $S_i^{2002} = \{no\ complete\ tumor\ resection, superficial\ tumor, prior\ TUR, no\ contraindication\ to\ cystectomy\}$, $S_j^{2004} = \{no\ complete\ tumor\ resection, superficial\ tumor, prior\ TUR, Ta, Grade\ I, no\ contraindication\ to\ cystectomy\}$, and $P_i^{2002} = P_j^{2004} = \{new\ TUR\}$.
5. *Recommendation refinement*: S_j is more specific than S_i and their action plans are similar (according to the plan similarity function). In this case, refinements occur at both clinical situation and action plan levels.
6. *New practice*: S_j is more specific than S_i , but their action plans are not similar.
7. *Unmatched recommendation*: R_i and R_j don't have anything in common.

```

1. if    $S_i = S_j$ 
2.   then if  $P_i = P_j$ 
3.     then No change
4.   else if  $Sim(P_i, P_j)$ 
5.     then Action plan refinement
6.   else New action plan
7.   else if  $Subsum(S_i, S_j)$ 
8.     then if  $P_i = P_j$ 
9.       then Condition refinement
10.    else if  $Sim(P_i, P_j)$ 
11.      then Recommendation refinement
12.    else New practice
13.    else Unmatched recommendation

```

Figure 1: Pseudo code for identifying evolution patterns from R_i to R_j

2.4. Comparison of recommendation sets

When comparing the two recommendation sets $\{R_{ij}^v\}$ and $\{R_{ij}^{v+1}\}$, which correspond respectively to the version v of a knowledge base and its following version $v+1$, each R_i^v is compared to each R_j^{v+1} with the above algorithm (Fig. 1) to determine which evolution pattern is relevant. At the end of the process, each recommendation R_j^{v+1} that always fell in the *Unmatched recommendation* case is considered *New recommendation*. Similarly, recommendations R_i^v that never matched any recommendation of the $v+1$ set are considered *Obsolete recommendations*.

3. Results

The method has been applied to bladder cancer. Textual 2002 CPGs and their 2004 revision were priorly formalized as decision trees. Each decision tree is structured to represent patient profiles clinicians may theoretically encounter as sequences of parameters (making the difference between clinical descriptors and therapeutic “coordinates” to locate the step the patient actually reached in the guideline-based care process) associated with the appropriate recommendations [10]. Structured KBs were then expanded as recommendations sets. The comparison of the 2 KBs is reported in table 1. In the 2004 version, many decision variables and actions were newly introduced, some were already in the prior version, while others were not used anymore. There were 577 recommendations in 2002 and 1,081 in 2004; only 47 were identical.

Table 1: Comparison of decision variables, actions and recommendations in 2002 and 2004 CPG versions

	2002-specific	Identical	2004-specific
Decision variables	8	32	23
Actions	9	27	22
Recommendations	530	47	1034

The detailed comparison between 2002 and 2004 recommendation sets was performed. For each decision variable and action, an abstract concept was associated to implement the *abst* function. Then, the algorithm described in the previous section was run. Table 2 reports the distribution of the distinct evolution patterns we identified in the 2004 update of the guidelines.

4. Discussion and conclusion

The evolution of bladder cancer CPGs between 2002 and 2004 mostly concerns the extension of CPG coverage. It is noticeable that only less than 5% of the recommendations remained unchanged and that two third of them were considered new. Thus, in this particular update, and within the proposed framework, less than one third of recommendations might be derived from the previous version. Compared to the initial version, nearly 60% of the 2002 recommendations were considered obsolete.

Table 2: Distribution of recommendation evolution patterns from the 2004 version point of view

Pattern	(n)	%	
Obsolete recommendation (2002)	339	58.8%	(2002-basis : total = 577)
No change	47	4.3%	(2004-basis : total = 1,081)
Action plan refinement	3	0.3%	
New action plan	40	3.7%	
Condition refinement	49	4.5%	
Recommendation refinement	39	3.6%	
New practice	180	16.7%	
New recommendation	723	66.9%	

The framework proposed to compare two successive versions of KBs can be used in knowledge-base-centric approaches. Indeed, it could provide health professional societies or national health agencies in charge of CPGs development with candidate documents produced by updated KBs. Another use could be to build “living KBs” where differences between two successive versions of KBs are easily recognizable and the types of the differences automatically identified.

5. References

- [1] Garg AX, Adhikari NKJ, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyenne J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA* 2005;293(10):1223–38.
- [2] Georg G, Séroussi B, and Bouaud J. Does GEM-encoding clinical practice guidelines improve the quality of knowledge bases? a study with the rule-based formalism. In: Musen M, ed, Proc AMIA 2003, Washington, DC. AMIA, November 2003:254–8.
- [3] Seyfang A, Martinez-Salvador B, Serban R, Wittenberg J, Miksch S, Marcos M, et al. Maintaining formal models of living guidelines efficiently. In: Bellazzi R, Abu-Hanna A, and Hunter J, eds, AIME 2007, (vol4594) of *Lecture Notes in Computer Science*, Amsterdam, NL. Springer, 2007:441–3.
- [4] Kaiser K and Miksch S. Formalizing ‘living guidelines’ using LASSIE: A multi-step information extraction method. In: Bellazzi R, Abu-Hanna A, and Hunter J, eds, AIME 2007, (vol4594) of *Lecture Notes in Computer Science*, Amsterdam, NL. Springer, 2007:401–10.
- [5] Shiffman RN, Michel G, Essaihi A, and Thornquist E. Bridging the guideline implementation gap: a systematic, document-centered approach to guideline implementation. *J Am Med Inform Assoc* 2004;11(5):418–26.
- [6] Shahar Y, Shalom E, and et al. AM. A distributed, collaborative, structuring model for a clinical-guideline digital-library. *J Am Med Inform Assoc* 2003:589–93.
- [7] Svatek V and Ruzicka M. Step-by-step formalisation of medical guideline content. *Int J Med Inf* 2003;70(2-3):329–35.
- [8] Votruba P, Miksch S, and Kosara R. Facilitating knowledge maintenance of clinical guidelines and protocols. In: MEDINFO, 2004:57–61.
- [9] Eriksson H, Tu SW, and Musen M. Semantic clinical guideline documents. *J Am Med Inform Assoc* 2005:236–40.
- [10] Séroussi B, Bouaud J, and Chatellier G. Guideline-based modeling of therapeutic strategies in the special case of chronic diseases. *Int J Med Inform* 2005;74(2-4):89–99.

Specification of business rules for the development of hospital alarm system: application to the pharmaceutical validation

Abdelali BOUSSADI ^a, Cedric BOUSQUET ^{a,b}, Brigitte SABATIER ^c, Isabelle COLOMBET ^{a,c} and Patrice DEGOULET ^{a,c}

^a *INSERM UMR_S 872, Eq 20, Medicine Faculty - Paris 5, France*

^b *University of Saint Etienne, Department of Public Health and Medical Informatics, St-Etienne, France.*

^c *Georges Pompidou European Hospital APHP, Paris, France.*

Abstract. Although clinical alarm systems are part of the knowledge management setting within healthcare organisations, modelling of business processes related to decision support and knowledge representation of decision rules are seldom described. We propose a customization of the Unified Process that takes into account user requirements for clinical alarm systems by introducing the Semantics of Business Vocabulary and Business Rules (SBVR). This methodology was applied to the design and implementation of a clinical alarm system for pharmaceutical validation at the European Hospital Georges Pompidou (HEGP). Rules were implemented using the IlogJRules Business Rule Management System. We produced 3 business rules patterns and 427 instances of rules. As SBVR is close to natural language, pharmacists were able to understand rules and participate to their design.

Keywords. Decision Support Systems, Clinical; Knowledge Representation (Computer); Knowledge Management; Decision Making, Computer-Assisted; Unified Process; Business rule.

Introduction

Many publications evaluating the errors of prescriptions justify the presence of alarm systems at the time of the prescription [1,2,3]. Computerized alarm systems implemented in Computerized Physician Order Entry (CPOE) systems have been proposed in an attempt to reduce these errors. However, their effectiveness depends not only on their intervention mode in the CPOE [4] but also on the compliance of prescribers to follow advice. In a study by Van der Sijs et al., prescribers override the alarms in 49% to 96% of the cases [3]. A second railing of prescriptions validation by hospital pharmacists, before drugs dispensation and administration can be assumed to be effective in preventing errors.

Since 1991, French legislation has mandated analysis and validation of drug prescriptions by pharmacists in hospitals [5]. A patient information system, integrating an electronic patient record and a CPOE (Dx-C@re, Medasys™) [6] is implemented

throughout our 800-beds hospital. The CPOE is used for every prescription of biology or radiology whereas it is used for drugs prescription only for 300 beds, the remaining prescription being done with paper and pencil. This organization allows the pharmacists to only validate computerized prescriptions, which already requires four full time pharmacists. The objective is to reach 650 beds in 2009. In order to maintain maximum safety, it is not envisaged to allow computerized drug prescription without pharmaceutical validation. It is not considered to increase the number of pharmacists for this task. The solution is to introduce a rule-based alarm system targeted to both the physician at the time of prescription and the pharmacist at the time of validation.

DoseChecker [7] is one of the rare expert systems introduced as an alarm system for pharmacists. In this study and in the majority of the studies published on the alarm systems there are no details on the methodology that led to the implementation of the system. To ensure that the alarm systems can deliver “good knowledge to the right people, in the right form and at the right time”, the designers of alarm systems must take into account two important criteria, which represent two real challenges [8].

First the developer needs to identify and model business processes (actors and business scenarios) in which the alarm system works. Several authors proposed an approach based on the object paradigm and more precisely on UML (Unified Modelling Language) for the modelling and design in a couple of decision support systems [9,10]. In these studies it is not specified if the use of the UML formalism is framed by a design method which takes into account the entire software life cycle (from the feasibility study to the maintenance). The authors of UML recommend the use of the Unified Process to cover all the steps related to the life cycle of the application.

Second the developer needs to model the knowledge associated with the decision rules with an appropriate language. Several languages have been proposed for the modelling of clinical practice guidelines [11]. However these languages present unnecessary complexity because they allow to represent specific features of clinical practice guidelines which are not part of rule based systems. The Arden syntax presents many limits: the data model is not object oriented, it is atomic and does not allow expressing the concept of time [12]. GELLO (object-oriented query and expression language) addresses the limits of Arden syntax, but it is based on a language intended for designers which involves a complexity in the writing and the reading of the business rules [12]. Ideally, the formalism selected should be articulated with the chosen design process and avoid complexity of writing or reading for the end user. It is the case of SBVR (Semantics of Business Vocabulary and Business Rules) [13].

Assuming that the same kind of rules can support alarms for prescribers and pharmacists, we concentrate on the modelling of the business processes of the pharmaceutical validation in the context of the Pompidou hospital central pharmacy. We aim at testing two research hypotheses (1) The Unified Process and UML notation are adaptable to allow business rules identification starting from business processes modelling intended for alarm system design, (2) SBVR allows the modelling of knowledge necessary to the writing of business rules. To examine these hypotheses we used Unified Process and UML for the modelling of the business processes. We used SBVR for modelling business rules in our alarm system for pharmaceutical validation. We implemented these business rules through a Business rule management system (ILOG JRules®) and assessed the value of Unified Process, UML and SBVR in our application context.

1. Materiel and method

Our method is articulated around the various phases and activities of Unified Process. We propose and justify an adaptation of Unified Process to our business and applicative context. The Unified Process is a development and design software process. Unified Process manages the design process according to several phases: Inception phase, Elaboration phase, Construction phase and Transition phase [14]. Each phase is composed of several activities (requirements, analysis...) more or less significant according to the applicative context and the phase in which the project is. We describe in what follows only the adaptations that we bring in the Elaboration and Construction phases.

1.1. Elaboration phase: Formulation and analysis of needs

In this phase we focus our attention on the requirements and analysis activities. Main goal of the requirements is the development of a model of the system to be built. The redaction of the system use cases constitutes an excellent way to proceed with the creation of this model.

The specification of the business rules and their introduction into the design cycle are not described in Unified Process. It is thus necessary to introduce a new stage which supplements the activity of formulation of the needs in the Elaboration phase as recommended by Nijpels [15] and Ambler [16]. This stage begins by identifying what are the business rules that have the greatest impact on the work of pharmacists in the context of the pharmaceutical validation. Then we establish outlines of business rules called *rules written pattern* according to Ambler's terminology.

- The first approach consists in using the data of the Pompidou Hospital Information System to evaluate the different clinical situations at risk at the time of the prescription and measure their frequency.
- The second approach is to consult the pharmacists in charge of the validation to identify the most significant business rules in the context of their activity.

1.2. Construction phase: design, implementation and testing

We associate for each use case a "use case realization – design". That passes by the identification of design classes formalized by the UML class diagram. The design class diagram related to one of *the rules written pattern* corresponds to our Business Object Model (BOM). The BOM corresponds to the UML business classes resulting from the construction phase of our design method. To instantiate the corresponding rules in this BOM, it is necessary to identify the relevant relations between classes and to name them. Then, by using the mapping rules between SBVR syntax and the BOM [17,18], we instantiate a business rule. For example class noun of the BOM corresponds to: noun concept in SBVR syntax and class name relation of the BOM corresponds to: <Role1>Verb<Role2> 'Fact type' in SBVR syntax. We repeat this procedure for each drug pertaining to the list of the drugs established with the pharmacists for each *rules written pattern* BRx by taking into account every related active ingredient. This produces all instances of rules.

The business rules of the alarm system are managed by a Business Rule Management System from ILOG: JRules. The implementation of our business rules

outcome of the design stage in JRules starts with the implementation of the eXecution Object Model (XOM) corresponding to our class diagram. The XOM is the model from which the business rules are implemented. The XOM maps a Java class to each class of the BOM and a Java attribute to each BOM attribute. The XOM can be built from: Compiled Java Classes (called: Java XOM), XML Diagram or Web services (called: Dynamic XOM). We generate one test case for each *rules written pattern* with a wizard provided by JRules. We edit the corresponding code source in order to enter typical conditions of use such as name of the drug, age of the patient, potassium level in input. The test case is successful when the rule is fired for abnormal conditions such as inappropriate potassium level or glomerular filtration rate.

2. Results

The results which we present follow the sequences of activity and the phases previously defined. To illustrate the application of our method we describe the successive stages of our method in one use case: 'To validate prescriptions'.

2.1. Business use case model and system use case model

We identified in consultation with the pharmacists in charge of the validation 7 business actors and two business use cases. The business use case "To validate the prescriptions" comprises one main scenario and 14 alternative scenarios. By taking into account the evaluations of the pharmacists in charge of the validation during various iterations, we identified in this stage three system use cases and three system actors related to the business use case "To validate the prescriptions" and corresponding to the three *rules written pattern BRx*.

The first *rules written pattern BR1*: 'control of the hyperkalemic drug prescriptions' was identified according to the first approach. The number of hyperkalemic drug prescription in patients with potassium level ≥ 5 mEq/L was 88 (1.2% for potassium supplementation) in one year. The second *rules written pattern BR2*: 'the adaptation of drugs dosage according to the Glomerular Filtration Rate and the third *rules written pattern BR3*: 'the adaptation of anticoagulant drugs dosage according to the International Normalized Ratio (INR) and the anti-Xa activity' were identified according to the second approach.

2.2. SBVR business rules

The BOM in Figure 1 corresponds to our patterns of business rules BRx. The mapping rules between SBVR syntax and the BOM allowed us to instantiate our business rules in their textual form (Figure 2).

- 247 business rules for the pattern of rules BR1.
- 156 business rules for the pattern of rules BR2.
- 24 business rules for the pattern of rules BR3.

Test cases corresponding to each *rules written pattern* were successful. For example we selected the following values for the business rule in figure 2:

- The attribute 'Speciality' of the class 'Drug' gets 'Heparin sodium ' ,
- The attribute 'Wording' of the class 'Route' gets 'Continuously ' ,

- The attribute 'clinical situation' of the class 'validation data' gets 'deep vein thrombosis',
- The attribute 'TCA' of the class 'patient' gets '2',
- The attribute 'anti Xa activity' of the class 'patient' gets '0.45'.

JRules triggers the corresponding rule and assigns to the attribute 'Wording' of the class 'Message' the value 'Valid prescription' that is displayed by the system to the user.

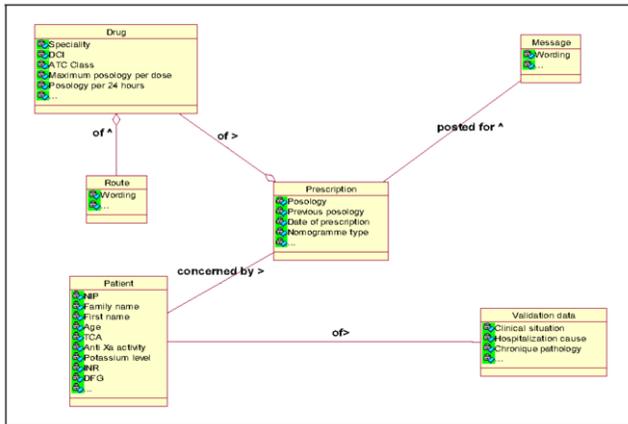


Figure 1. The BOM corresponding to the rule pattern BR3

if the speciality of the drug of the prescription is Heparin sodium and the wording of the route of the drug of the prescription is continuously and the clinical situation of the validation data of the patient concerned by the prescription is deep vein thrombosis and (the TCA of the patient is equal to 2 times the witness or anti Xa activity is between 0.4 and 0.6) then the wording of the message posted for the prescription is valid prescription

Figure 2. Rule instance of the BR3 rule pattern following the SBVR syntax

3. Discussion and conclusion

The choice of SBVR as a specification formalism of the business rules enabled us to write business rules with a vocabulary which mirrors the natural language. SBVR allows to instantiate rules starting with a design class diagram. That implies that there is no dichotomy between the adaptation of Unified Process which we propose and the use of SBVR.

The design class diagram which we highlighted in the construction phase is the result of only two iterations. Our class diagram must be supplemented with an existing object oriented model for the modeling of the drugs prescriptions [19]. The rules which we implemented in our system are intended for the dosage adaptation according to biology results. The extension of the system to other categories of rules in particular those which address contraindications (drug - disease interactions) is a significant perspective [20]. In order to carry out more precise tests, the rules implemented in JRules should be integrated in the Pompidou Hospital Information System to exploit the patient's data of DxCare. An additional desirable data-processing development is to couple the rule system with a drug knowledge base.

All the studies which we listed within the background of our work do not give and do not recommend precise method for the implementation and the design of rule-based

alarm systems. This work represents a first stage in the establishment of a design methodology for the implementation of alarm systems.

Acknowledgments

We acknowledge the ILOG Compagny for providing an evaluation version of ILOG JRules. Requirements for the rule system were written thanks to Dr Philippe Guillaume and Dr Thibaut Caruba from the pharmacy of the Pitié-Salpêtrière hospital.

References

- [1] Moore N, Lecointre D, Noblet C, Mabilille M. Frequency and cost of serious adverse drug reactions in a department of general medicine. *Br J Clin Pharmacol*. 1998;45(3):301-8.
- [2] Oliven A, Michalaké I, Zalman D, Dorman E, Yeshurun D, Odeh M. Prevention of prescription errors by computerized, on-line surveillance of drug order entry. *Int J Med Inform*. 2005;74(5):377-86.
- [3] Van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. *J Am Med Inform Assoc*. 2006;13(2):138-47.
- [4] Kawamoto K, Houlihan CA, Balas EA, Lobach DF Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ*. 2005
- [5] Vernardet S, Bossaert S, Livrozet A, Pont E, Charpiat B. Pharmaceutical analysis and validation of hospital prescriptions, a 5-year study. *Presse Med*. 2005;34(14):990-6.
- [6] Degoulet P, Marin L, Lavril M, Le Bozec C, Delbecke E, Meaux JJ, Rose L. The HEGP component-based clinical information system. *Int J Med Inform*. 2003 ;69(2-3):115-26.
- [7] McMullin ST, Reichley RM, Kahn MG, Dunagan WC, Bailey TC. Automated system for identifying potential dosage problems at a large university hospital. *Am J Health Syst Pharm*. 1997 1;54(5):545-9.
- [8] Peleg M, Tu S. Decision support, knowledge representation and management in medicine. *Methods Inf Med*. 2006;45 Suppl 1:72-80.
- [9] LeBozec C, Jaulent MC, Zapletal E, Degoulet P. Unified modeling language and design of a case-based retrieval system in medical imaging. *Proc AMIA Symp*. 1998;:887-91.
- [10] T. Knape, L. Hedermana, VP. Wadea, M. Gargana, C. Harrisa, Y. Rahmanb, A UML Approach to Process Modelling of Clinical Practice Guidelines for Enactment, *IOS Press*, 2003, 635-640.
- [11] Peleg M et al. Comparing computer-interpretable guideline models: a case-study approach. *J Am Med Inform Assoc*. 2003;10(1):52-68.
- [12] Sordo M, Boxwala AA, Ogunyemi O, Greenes RA. Description and status update on GELLO: a proposed standardized object-oriented expression language for clinical decision support. *Medinfo*. 2004;11(Pt 1):164-8.
- [13] "Anonyme" Semantics of Business Vocabulary and Business Rules Specification, Document de specification de l'OMG, 2006. <http://www.omg.org/docs/dtc/06-03-02.pdf> last accessed on 2007/05/02
- [14] In *The Unified Software Development Process*, eds Ivar Jacobson, Grady Booch, James Rumbaugh, Addison-Wesley Professional, 1999.
- [15] Nijpels R. Business Rules in Requirements Analysis <http://www.brcommunity.com/b259.php#fn1> last accessed on 2007/05/15
- [16] Ambler S. How to record the transactional logic behind UML object models <http://www.ddj.com/dept/architect/184414624?cid=Ambysoft> last accessed on 2007/05/15
- [17] "Anonyme" Business Semantics of Business Rules, *Business Rules Team Response to RFP: SBVR Submission bei/2005-08-01* Version 8. <http://www.omg.org/docs/bei/05-08-01.pdf> last accessed on 2007/05/02
- [18] Schacher M. Moving from Semaman Row 2 to Zachman Row 3 Business Rules from an SBVR and an xUML Perspective, 2006, KnowGravity.
- [19] Sene B, Venot A, de Zegher I, Milstein C, Errore S, de Rosis F, Strauch G. A general model of drug prescription. *Methods Inf Med*. 1995;34(4):310-7.
- [20] Kuperman GJ, Bobb A, Payne TH, Avery AJ, Gandhi TK, Burns G, Classen DC, Bates DW. Medication-related clinical decision support in computerized provider order entry systems: a review. *J Am Med Inform Assoc*. 2007;14(1):29-40.

Formalization of Clinical Practice Guidelines

David BUCHTELA, Jan PELEŠKA, Arnošt VESELÝ, Jana ZVÁROVÁ,
Miroslav ZVOLSKÝ

EuroMISE Center and Center for Biomedical Informatics, Department of Medical Informatics, Institute of Computer Science AS CR, v.v.i., Pod Vodarenskou vezi 2, 182 07 Prague 8, Czech Republic

Abstract: Clinical practice guidelines are textual recommendations based on the consensus of medical experts with the aim to solve diagnostic and therapeutic problems. For more advanced use in real medical applications it is necessary to find out mathematical models of physicians' decision-making processes. The acquisition of a formal model from text-based guidelines is a crucial point for development of decision support systems. We introduce a system for formalization and presentation of medical knowledge contained in clinical practice guidelines where knowledge formalization is based on the GLIF model.

Keywords: biomedical informatics, formalization of medical knowledge, clinical practice guidelines, GLIF model

Introduction

Clinical practice guidelines (CPG) are developed as textual recommendations by groups of medical experts on topics selected by a local scientific authority (expert medical society) or by a national health institution. Their goal is to improve the quality of medical care and to achieve treatment standardization. The outcoming text is created for a specific group of physicians, other health professionals or patients. For computer implementation and processing it is necessary to have guidelines structured explicitly [1]. CPGs are published mostly in printed versions as a paper or as a part of a paper in scientific journals. Nowadays, CPGs are published mostly on the Internet, too. Formalization of guidelines by means of a general GLIF model presents a suitable additional educational tool for their easier knowledge implementation in comparison with the classical paper form [2, 3]. It also allows to use patients' data from an electronic health record (EHR) system. The formalization of CPGs can function as a feedback for authors to remove uncertainties and information inconsistencies in CPGs [4].

The aim was to create an easy-to-use system for converting medical knowledge included in text-based clinical practice guidelines into a formal model and save it in a widely applicable format (XML). The goal was to use it in developed applications for presentation, educational purposes and also in more complex systems for decision support in future.

1 Single purpose application

As the first step of using knowledge extracted from text-based guidelines we developed a web-based tool, which included the decision-making algorithm of the 1999 WHO/ISH Hypertension Guidelines. After a completion of a special form, physicians receive an automated assessment of a patient's cardiovascular risk based on three groups of items – risk factors, target organ damages and associated clinical conditions known from the guidelines paper form. All completed patient's items are used for a computation of his/her individual cardiovascular risk presented as a highlighted field of the known table from the guidelines paper form. When the drug treatment is recommended, a list of items for selecting the right drug treatment based on graded indications and contraindications can be filled in for an individual patient to obtain a list of recommended drug classes [5]. A significant disadvantage of this application was the necessity of a special application for every newly used guidelines and manual data entry for each individual processed patient.

2 GLIF based applications

2.1 GLIF model

For knowledge formalization in clinical practice guidelines we decided to use the universal and widely respected GLIF model. GLIF specifies an object-oriented model for guidelines representation and syntax for guidelines utilization in software systems as well as for their transfer. GLIF guidelines are mostly given as a flowchart representing a temporarily ordered sequence of steps. The nodes of the graph are guidelines steps and the edges represent continuation from one step to the other. The guidelines steps are: *action step*, *decision step*, *branch* and *synchronization steps*, and a *patient state step* [6], [7]. The decision step specifies several criteria of condition for each decision option. The *strict-out* criterion is evaluated at first. If the *strict-out* criterion is evaluated as true, the rest of criteria is not evaluated. This option is forbidden. In the opposite case the *strict-in* criterion is evaluated. If the *strict-in* criterion is false too, the *rule-in* and *rule-out* criteria are evaluated. The ranking of *rule-ins* and *rule-outs* is left for the user who may use his or her clinical judgment or may develop their own ranking schemes.

The GLIF model is graphical and therefore it is necessary to code it in the XML form. Syntax for a guidelines describing language is a part of a guideline model specification. In a language form encoded guidelines consist of a sequence of guidelines steps. Some attributes of a guidelines step contain next guidelines steps. It enables sequential representation of a graph structure in the guidelines language [8]. GLIF also enables a connection of particular guidelines in one complex sequence suitable for usage in clinical practice. It is shown in the GLIF model of formalized 2003 European Guidelines on Cardiovascular Disease Prevention in Clinical Practice and those of associated diseases, which offers to physicians a system for a decision support and checks their decision algorithms in comparison with those of guidelines [9].

2.2 GLIF model construction

A GLIF model construction and implementation of text guidelines are not easy. The whole process can be divided into several stages (see Fig. 1).

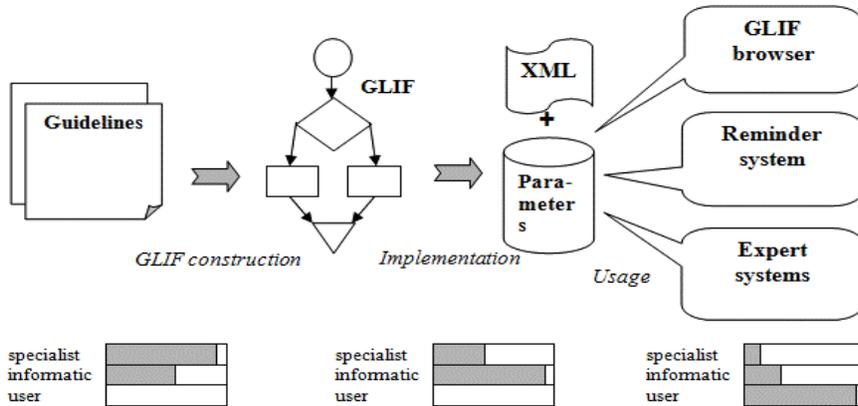


Figure 1: Process of GLIF model construction, implementation and usage

In the stage of GLIF model construction from text guidelines, it is important to find a logical and process structure of guidelines, all fundamental parameters and their interrelationships. Cooperation of an informatician and a medical specialist (the author of text guidelines is preferred) is more effective. The result of this cooperation is a graphic GLIF model that corresponds with the text guidelines. The construction stage is the most important and difficult stage. The specialist can use the developed user-friendly GLIF editor in the construction stage (see Fig. 2).

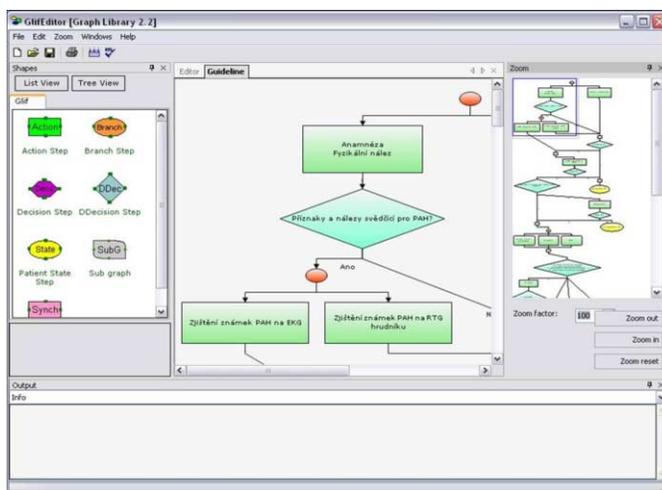


Figure 2: GLIF editor

In contrast to systems that attempt to automate the conversion of text-based guidelines into a formal model usable in the health information systems [10, 11, 12], the GLIF editor offers the easy way of this model creation, which could be done by one

of guidelines authors. The usability of the GLIF editor was tested on a number of guidelines created by the Czech Society of Cardiology.

In the stage of GLIF model implementation, the graphic model of guidelines is coded into XML. Moreover, a list of basic and derived parameters is created. Basic parameters represent measurable values directly. Derived parameters are obtained in an arithmetical, logical or logically-arithmetical operation above basic parameters. The cooperation of IT specialists and medical experts also plays an important role in the creation of a basic and derived parameters list. The result is a data model that serves as an interface between the GLIF model and real input data stored in EHR (*Electronic Health Record*). It is important to pay attention to the definition of all criteria of conditions (*strict-in, strict-out, rule-in, rule-out*) for each decision option.

2.3 Applications using GLIF model of CPGs

The *educational version* of the CPG presentation system was designed to be able to combine the HTML text, GLIF model graphs and specialized presentation modules where the standard GLIF model presentation would be too complex. Guidelines are shown in any Java enabled Internet browser (e.g. Internet Explorer with installed Java Runtime Environment). The *Glifview* is a GLIF model browser that can present any formalized medical guidelines in a user-friendly manner (see Fig. 3). The educational version of CPG using the GLIF model leads a physician through the decision tree usually with yes-no alternatives in dependence on the physician's knowledge of patient's data. If a value of some variable is not available, the physician can continue and simulate both possible alternatives in the concerned decision step.

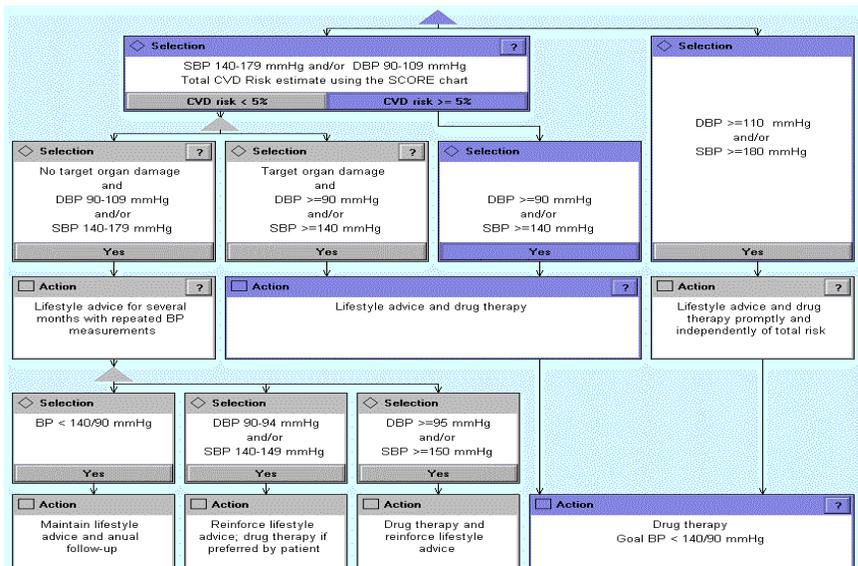


Figure 3: Glifview example

The *processing version* of CPG enables physicians to put patient's data into the model directly. There is a list of all the variables used, the values of which are to be filled in by a physician. Compared to the educational system the browser goes

automatically through the GLIF model graph evaluating conditions of decision steps. If some conditions cannot be evaluated, as the needed data items are not available, the browser stops and highlights the branch from the root to the current step. Thus it can serve as a reminder of missing data necessary for a correct decision. Then the user inputs missing data (or simulates data) to the browser manually to be able to continue in visualization.

The Medical Knowledge Representation System (MEKRES) for diagnostic and treatment support is developed. The system will automatically offer to participants (patient, general practitioner, operator, ...) relevant formalized guidelines on the basis of acquired patients' data (monitored or saved in EHR). The selection of personalized guidelines will be provided according to key attributes of CPGs in cooperation with generally accepted standards, nomenclature and classification systems (see Fig. 4).

Medical knowledge and guidelines will be modelled by extended GLIF models (Knowledge Representation Model - KREM). The key attributes (code of a disease and others) will be added to the GLIF model for an easier identification and searching for a formalized guideline.

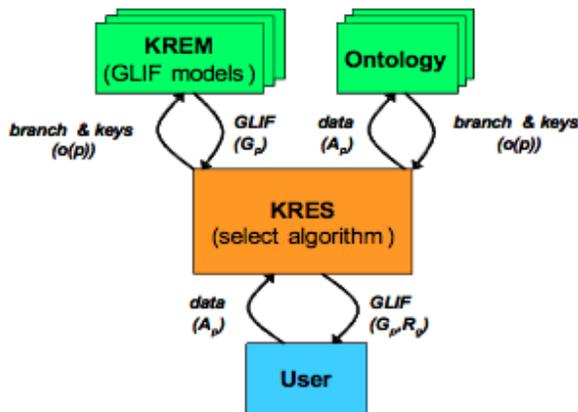


Figure 4: Knowledge Representation System

3 Conclusions

Several applications for formalization and presentation of guidelines were developed in the EuroMISE Center. The GLIF editor enables an easy way to create GLIF models of text-based CPGs. For educational purposes and usage in clinical practice the Glifview component and processing guidelines system were created. Several projects with a similar purpose have been developed all around the world recently [13, 14, 15]. The creation of a proper system of formalisation and presentation of CPGs is a tendency of using the generally accepted format (GLIF) and widening it for our needs in the context with our future plans.

These applications are employed in the test mode in the EuroMISE Center. In cooperation with the Czech Society of Cardiology and the Society of General Practice some of their CPGs were formalized (e.g. CPG for the Management of Arterial Hypertension, Guidelines for Treating Tobacco Dependence, Guidelines for Diagnosis

and Treatment of Acute Myocardial Infarction, Guidelines for Diagnosis and Treatment of Pulmonary Embolism). These applications also serve as educational tools and they are prepared for use with EHR.

Acknowledgement

The work was partially supported by the project 1M06014 of the Ministry of Education, Youth and Sports of the Czech Republic and by the research institutional plan AV0Z10300504 of the Institute of Computer Science AS CR, v.v.i.

References

- [1] Vesely A., Zvarova J., Peleska J., Buchtela D., Anger Z.: Medical guidelines presentation and comparing with Electronic Health Record. *International Journal of Medical Informatics*, Vol. 75, Issues 3-4, 2006, pages 240-245.
- [2] Peleska J., Anger Z., Aschermann M., Buchtela D., Tomeckova M., Vesely A., Zvarova J.: Formalization of medical guidelines, in: *Proceedings of the International Joint Meeting EuroMISE Prague*, 2004.
- [3] Quaglini S., Ciccarese P.: Models for guidelines representation, *Neurological Sciences Journal*, Springer, 2007
- [4] Peleska J., Anger Z., Buchtela D., Sebesta K., Tomeckova M., Vesely A., Zvara K., Zvarova J.: Formalization of medical guidelines, *European Journal for Biomedical Informatics*, 2004.
- [5] Peleska J., Zvara K., Vesely A., Zvarova J., Tomeckova M., Hanzlicek P.: Development of electronic form of the 1999 WHO/ISH hypertension guidelines, *Stud Health Technol Inform.* 2002;90:268-71.
- [6] Ohno-Machado L., Gennari J. H., Murphy S., Jain N., Tu S., W., Oliver D., et al.: The GuideLine Interchange Format: A model for representing guidelines, *Journal of the American Medical Informatics Association* 1998, 5(4), pp. 357-372.
- [7] Ash N., Bernstam E., Greenes R.A., Lacson R., Mork P., Shortliffe E.H.: Guideline Interchange Format 3.5 Technical Specification, online - 12.12.2002, In: <http://smi-web.stanford.edu/projects/internet-web/guidelines/>
- [8] Buchtela D., Peleska J., Vesely A., Zvarova J.: Presentation of Medical Guidelines on a Computer, Transformation of Healthcare with Information Technologies, Ed.: Zielinski K., Duplaga M., Ingram D., IOS Press, Amsterdam, 2004, pp.166-171, ISBN 1-58603-438-3.
- [9] Anger Z., Buchtela D., Peleska J., Tomeckova M., Vesely A., Zvarova J.: Formalized 2003 European Guidelines on Cardiovascular Disease Prevention in Clinical Practice, <http://www.euromise.cz/health/recommendation/cvd2004.html>[10] Ruzicka M., Svatek V.: Markup Based Analysis of Narrative Guidelines with the Stepper Tool, *Computer-based Support for Clinical Guidelines and Protocols*, IOS Press 2004, pp. 132-136
- [11] Gersende G., Cavazza M.: Integrating Document-Based and Knowledge-Based Models for Clinical Guidelines Analysis, *Artificial Intelligence in Medicine*, Volume 4594, 2007
- [12] Serban R., ten Teije A.: Formalization of medical guidelines exploiting medical thesauri, *Proceedings of the European Federation of Medical Informatics Special Topic Conference 2006*, IOS Press
- [13] Tu S. W., Campbell J., Musen M. A.: The SAGE Guideline Modeling: Motivation and Methodology, *Computer-based Support for Clinical Guidelines and Protocols*, IOS Press 2004, pp. 162-166
- [14] Laleci G. B., Dogac A., Akcay B., Olduz M., Yuksel M., Orhan U., Tasyurt I., Sen T., Kabak Y., Namli T.: SAPHIRE: A semantic Web service based Clinical guideline deployment infrastructure exploiting IHE XDS, *Exploiting the Knowledge Economy: Issues, Applications, Case Studies*, Paul Cunningham and Miriam Cunningham (Eds), IOS Press, 2006 Amsterdam, ISBN: 1-58603-682-3, pp. 917-923
- [15] Terenziani P., Montani S., Bottrighi A., Torchio M., Molino G., Correndo G.: The GLARE Approach to Clinical Guidelines: Main Features, *Computer-based Support for Clinical Guidelines and Protocols*, IOS Press 2004, pp. 162-166

A Simple Method for Heuristic Modeling of Expert Knowledge in Chronic Disease: Identification of Prognostic Subgroups in Rheumatology

Örjan DAHLSTRÖM^{a, b, 1}, Toomas TIMPKA^{c, d}, Ursula HASS^c,
Thomas SKOGH^e and Ingrid THYBERG^c

a The Swedish Institute for Disability Research, Linköping and Örebro University, Sweden

b Department of Behavioural Sciences and Learning, Linköping University, Sweden

c Department of Health and Society, Linköping University, Sweden

d Department of Computer and Information Science, Linköping University, Sweden

e Department of Molecular and Clinical Medicine, Linköping University, Sweden

Abstract. Identification of prognostic subgroups is of key clinical interest at the early stages of chronic disease. The aim of this study is to examine whether representation of physicians' expert knowledge in a simple heuristic model can improve data mining methods in prognostic assessments of patients with rheumatoid arthritis (RA). Five rheumatology consultants' experiences of clinical data patterns among RA patients, as distinguished from healthy reference populations, were formally represented in a simple heuristic model. The model was used in K-mean-clustering to determine prognostic subgroups. Cross-sectional validation using physician's global assessment scores indicated that the simple heuristic model performed better than crude data made in identification of prognostic subgroups of RA patients. A simple heuristic model of experts' knowledge was found useful for semi-automatic data mining in the chronic disease setting. Further studies using categorical baseline data and prospective outcome variables are warranted and will be examined in the Swedish TIRA-program.

Keywords. Knowledge engineering, Clinical Decision Support Systems, Semi-automated Data Mining, Rheumatoid Arthritis, Mathematical models in medicine.

Introduction

Chronic diseases, such as cardiovascular disease, diabetes, and rheumatic diseases constitute a large part of the disease burden in western countries [1]. Early identification of patients with chronic disease at risk of progression is essential, especially when early intervention can induce remission or retard the disease process.

¹ Corresponding Author: Örjan Dahlström, Department of Behavioural Sciences and Learning, SE-581 83 Linköping, Sweden. E-mail: orjan.dahlstrom@liu.se.

The prevalence of rheumatoid arthritis (RA) in an adult Swedish population is about 5-7 cases per 1000 individuals [2] and the yearly incidence is 24/100,000 [3]. Identifying subgroups of patients with a less favorable prognosis at an early stage is therefore both a key clinical and public health priority.

Much of clinical decision support research has addressed the establishment of a clinical diagnosis during a short-term care episode. This approach has often been developed using comparisons between physicians' performances as a baseline, implying a replacement of the decision-maker rather than integration of the decision support system as a component in the actual clinical care environment. Such an approach has less relevance in the chronic disease context, where the diagnosis is already known. A more appropriate strategy here is that the decision support supplies the decision-maker with contextual information at different stages in the clinical process. This strategy is compatible with theories of distributed cognition [4], in which both human and machine agents are considered as being integral parts of a functional cognitive system, that recently have been applied in clinical environments [5].

A specific type of decision support is assistance with knowledge discovery in databases (KDD) [6]. Integration of background knowledge into discovery methods can improve the quality of these processes. This tactic has been used for, e.g. identification of new clinically interesting subgroups [7]. However, putting the medical expert in the center of the KDD process, described as active mining, can be time-consuming and also expensive. Therefore, semi-automated active mining has been introduced. This method has shown to be promising with regard to automated discovery from an incomplete data set in early detection of patient groups at risk for coronary heart disease [8].

The aim of this study is to examine whether representation of physicians' expert knowledge in a simple heuristic model can be used to improve data mining methods in prognostic assessments of RA patients. The starting point for the heuristic modeling is that the clinical test values for patients with a specific chronic disease differ from those of healthy reference populations.

1. Materials and Methods

The research was performed in three steps. In a *knowledge engineering step*, physicians' experiences from the clinical use of four variables included in the evaluation of RA patients were elicited. In the second step, these data were analyzed and represented in a model that was implemented in an algorithm for the *determination of prognostic groups*. For comparison, the same algorithm was also used to develop groups based on the crude data set. In the final step, *prognostic group validation*, the model was validated by comparing physicians' global assessment of disease activity scores (PGA (scores 0-4, where 0 corresponds to no activity and 4 represents high activity)) for the patient groups identified with and without the model.

Knowledge Engineering

The knowledge engineering was based on data from a Swedish 'early arthritis' database [9]. This resource has been built in a prospective multi-center study (TIRA) where 320 patients diagnosed with RA were diagnosed early and included during 1996 - 1998. The patients fulfilled $\geq 4/7$ RA classification criteria (95%), as defined by the

American College of Rheumatology 1987, or exhibited morning stiffness ≥ 60 minutes, symmetrical arthritis, and arthritis of small joints (5%). Although several prognostic indicators can be suggested, the interval-/ratio variables in the database were especially interesting for this study, because they could be used in the standard method of K-mean clustering. The set of variables used here included the erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), the ‘number of swollen joints’ and the ‘number of tender joints’.

Five consultants in rheumatology were interviewed. The interviews were held in a semi-structured manner and aimed at investigating the physicians’ perspectives on important variables when making prognostic judgments about patients with RA. The consultants had the opinion that the variables ‘number of swollen joints’ and ‘number of tender joints’ were not sufficient stand-alone indicators of the patient’s status. These were therefore omitted from further analysis. ESR and CRP, however, were both considered to be important and relevant indicators of patients’ status. We wanted the modeling of their expert knowledge to be as simple as possible to interpret and therefore used the notions values. We therefore modeled their views on ESR and CRP data by using the values they defined as low, elevated, and definitely high as limits for categorizations of ESR and CRP. The consultants’ perspectives on what are considered as low values were almost identical ($ESR \leq 12$ mm for men and $ESR \leq 20$ mm for women, $CRP \leq 10$ mg/L for both men and women), with the exception of some minor differences regarding low values for ESR. Opinions of what are to be considered as elevated or definitely high values differed only slightly (an elevated value is roughly $ESR = 60$ mm and $CRP = 70$ mg/L, while $ESR = 100$ mm and $CRP = 100$ mg/L are definitely high values).

To give ESR and CRP equal influence we transformed the ESR and CRP values to the interval $[0, 1]$. We estimated 90% of the values to be represented by values spanning from low to elevated (0, .9) by a linear transformation. Values ranging from elevated to definitely high were transformed into the interval $[.9, 1]$ in a decreasing manner using Eq. (1). Values considered as low were transformed to 0 and values considered as definitely high were transformed to 1.

$$f(x) = p + (1 - p) \cdot \left[1 - \left(1 - \frac{x - x_1}{x_2 - x_1} \right)^{\frac{\frac{p}{x_1 - x_0} (x_2 - x_1)}{1 - p}} \right] \tag{1}$$

In Eq. (1). x is the variable for which the consultants’ experience is being modeled (ESR or CRP), f is the modeled value, x_0 is the ‘lower limit value’ for which all lower values are considered as low, x_1 is the ‘first upper limit value’, the point at which higher values are considered as elevated, x_2 is the ‘final upper limit value’ above which all higher values are considered definitely high, p is the width of the interval which $[x_0, x_1]$ is transformed to.

The ‘limit values’ used in Eq. (1) for ESR and CRP in the knowledge engineering step are shown in Table 1.

Table 1. Values for model limits x_0 , x_1 , x_2 , and p determined in the knowledge engineering step when modeling consultants' views of erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP).

Variable	x_0	x_1	x_2	p
ESR (men)	12	60	100	.90
ESR (women)	20	60	100	.90
CRP (men and women)	10	70	100	.90

Determination of Prognostic Groups

To find different subgroups of patients, the method of K-mean clustering, was used on crude data, and then with two kinds of transformed data:

- Model A – Using the transformed data.
- Model B – Using the transformed data but with a predefined group of patients with values below the lower limit values (x_0) for both ESR and CRP.

To keep the results on a clinically comprehensive and not too complex level, the number of prognostic subgroups was limited to a maximum of four groups. The case with two subgroups was found trivial. Further cluster analyses were therefore restricted to three and four subgroups.

Prognostic Group Validation

In each set of the cluster analyses, the patients in the prognostic groups were compared with one another with respect to PGA. Subgroups, identified when the models were used, were compared with the results based on crude data to establish which method that best identified prognostic subgroups with respect to PGA using 95% confidence interval.

2. Results

Determination of Prognostic Groups

Running K-mean-clustering using three clusters resulted in a 'low value group' (low values for both ESR and CRP), a 'medium value group' (medium values for both ESR and CRP), and a 'high value group' (high values for both ESR and CRP). This is what could have been expected since there is a positive correlation (.66) between ESR and CRP. The same is true for the transformed values where the correlation between the modeled values for ESR and CRP are positive, varying between .66 and .68.

Running K-mean-clustering using four clusters made the situation more complex. As in the case of three clusters, a 'low value group' and a 'high value group' appeared. There were also two 'medium value groups'. Despite minor differences in sizes and medium values for the clusters in the different model(s), the results were quite similar in this respect.

Table 2. Validation of derived models A and B against crude data using physicians' global assessment scores (PGA) as proxy for patient outcome. Mean values (95% confidence interval) are displayed for each cluster (three and four clusters, respectively).

3 Clusters	Crude data		Model A		Model B	
	N	PGA	N	PGA	N	PGA
1 (low)	188	1.89 (1.78-2.01)	154	1.84 (1.71-1.96)	59	1.56 (1.37-1.76)
2 (medium)	70	2.10 (1.93-2.27)	72	2.04 (1.87-2.21)	149	1.99 (1.87-2.11)
3 (high)	21	2.33 (1.89-2.77)	53	2.30 (2.08-2.53)	71	2.30 (2.12-2.48)
4 Clusters						
1 (low)	171	1.88 (1.76-2.00)	146	1.82 (1.69-1.95)	59	1.56 (1.37-1.76)
2 (medium)	36	2.19 (1.98-2.41)	31	2.16 (1.95-2.38)	110	2.00 (1.86-2.14)
3 (medium)	51	2.00 (1.77-2.23)	62	2.05 (1.86-2.23)	58	2.07 (1.87-2.27)
4 (high)	21	2.33 (1.89-2.77)	40	2.30 (2.02-2.58)	52	2.31 (2.08-2.54)

Prognostic Group Validation

In the case of three clusters, the three patient subgroups identified using model B on the basis of ESR and CRP differed significantly with regard to PGA scores (Table 2). Model A identified one significant difference. In comparison, the analyses based on the crude data identified no significant differences between any groups.

In the case of four clusters, no significant differences were found in the crude data model with regard to PGA. Model A and Model B, however, both allowed the identification of significantly different groups based on ESR and CRP (Model A: 1 significant difference, Model B: 3 significant differences).

3. Discussion

A knowledge engineering approach where heuristic expert knowledge was used to transform data by a simple procedure showed to be more efficient for identifying prognostic subgroups than K-mean clustering of crude data. The resulting subgroups passed the 'objective' gateway, which means that they included significant differences when using basic statistical criteria for cross-sectional evaluation, compared to the analysis of crude data [8]. The PGA score used for the evaluation is the variable in the TIRA database that was the closest proxy to long-term clinical outcomes. Passing the next, 'subjective', gateway will require that benefit is perceived when the procedure is used in a clinical action scenario.

Our approach to identifying prognostic subgroups differs from most other approaches in subgroup discovery [10]. These procedures mainly address the problem of large search spaces using constraint knowledge (restricting the search space), pattern knowledge (focusing the search) or ontological knowledge (applying weights to attributes) [7]. In other words, these methods are relevant mainly for larger datasets. Engineering knowledge in such settings aims usually to exclude uninteresting parts of the database. However, our main interest was not to mine data in large datasets, but to

examine heuristic modeling of expert knowledge in a form usable in standard data mining methods. We employed cluster analysis on a medium-sized dataset ($n = 320$), but our method can also be used in other methods and on larger datasets. For instance, our knowledge engineering modeling of expert knowledge is possible to integrate into other subgroup discovery approaches [7, 8] and into other general KDD approaches.

A main reason for using cluster analysis as the standard method for subgroup discovery is that understandability and interpretability of the generated models are of prime importance when viewing the knowledge discovery from the expert's point of view. Results from cluster analysis are relatively simple to interpret and trace compared to the output from many other data mining approaches, e.g. support vector machines. The use of cluster analysis is an attempt to 'open the black box' of computations in data mining [11] for the clinician by choosing a method that is easy to understand. In the clinical context of chronic disease management, clinicians will be able to use our results in semi-automated data mining as part of the KDD processes. HEARTFAID is a comprehensive program that aims to improve the prognostic processes in the field of chronic cardiovascular disease [12]. The program has developed a similar architecture for an information infrastructure that is under development in the TIRA program in rheumatology. Our simple heuristic model fits into both of these contexts.

An important limitation of the model in its present form is that it is applicable only to variables on the interval-/ratio scale. The next step in the development of the model in the context of semi-automated data mining will be to include the examination of categorical data. Further research involving prospective outcome variables is warranted.

References

- [1] A. Bakhai A. *The burden of coronary, cerebrovascular and peripheral arterial disease*. Pharmacoeconomics 22 (2004) (Suppl 4),11-8.
- [2] M. Simonsson, S. Bergman, L.T. Jacobsson, I.F. Petersson, B. Svensson. *The prevalence of rheumatoid arthritis in Sweden*. Scand J Rheumatol **28** (1999) (6),340-343.
- [3] M.K. Söderlin, O. Börjesson, H. Kautiainen, T. Skogh, M. Leirisalo-Repo, *Annual incidence of inflammatory joint diseases in a population based study in southern Sweden*. Ann Rheum Dis **61** (2002), 911-915.
- [4] E. Hutchins. *Cognition in the wild*. MIT Press, Cambridge (MA), 1995.
- [5] M. Bång, T. Timpka. *Cognitive tools in medical teamwork: the spatial arrangement of patient records*. Methods Inf Med 42 (2003), 331-336.
- [6] L.A. Kurgan, P. Musilek, *A survey of knowledge discovery and data mining process models*. Knowledge Engineering Review 21 (2006),1-24.
- [7] M. Atzmueller, F. Puppe, H.-B. Buscher, *Exploiting background knowledge for knowledge-intensive subgroup discovery*. In: Proceedings of IJCAI-05, IJCAI, New York, 2005, p. 647-652.
- [8] A. Silberschatz, A. Tuzhilin, *What makes patterns interesting in knowledge discovery systems*, IEEE Transactions on Knowledge and Data Engineering 8 (1996), 970-974.
- [9] I. Thyberg, U.A. Hass, U. Nordenskiöld, B. Gerdle, T. Skogh, *Activity limitation in rheumatoid arthritis correlates with reduced grip force regardless of sex: the Swedish TIRA project*. Arthritis Rheum 53 (2005), 886-896.
- [10] N. Lavrač, B. Cestnik, D. Gamberger, P. Flach. *Decision support through subgroup discovery: three case studies and the lessons learned*. Machine Learning 57 (2004) (3),115-43.
- [12] R. Amar, J. Stasko, *A knowledge task-based framework for design and evaluation of information visualizations*. Proceedings Infovis 2004, p. 143-149.
- [13] D. Conforti, D. Costanzo, F. Perticone, et al, *HEARTFAID: A knowledge based platform of services for supporting medical-clinical management of heart failure within elderly population*. Stud Health Technol Inform. 121 (2006), 108-125.

A Method for Indexing Biomedical Resources over the Internet

Guillermo DE LA CALLE ^a, Miguel GARCIA-REMESAL ^a and Victor MAOJO ^a

^a *Biomedical Informatics Group, Artificial Intelligence Lab., Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo s/n, 28660 Boadilla del Monte, Madrid, Spain*

Abstract. A large number of biomedical resources are publicly available over the Internet. This number grows every day. Biomedical researchers face the problem of locating, identifying and selecting the most appropriate resources according to their interests. Some resource indexes can be found in the Internet, but they only provide information and links related to resources created by the owner institution of each website. In this paper we propose a novel method for extracting information from the literature and create a Resourceome, i.e. an index of biomedical resources (databases, tools and services) in a semi-automatic way. In this approach we consider only the information provided by the abstracts of relevant papers in the area. Building a comprehensive resource index is the first step towards the development of new methodologies for the automatic or semi-automatic construction of complex biomedical workflows which allow combining several resources to obtain higher-level functionalities.

Keywords. Data Acquisition, Data Capture, Classification, Indexing

Introduction

Over the last years, there has been a proliferation of biomedical resources, including databases, tools, and services. Many of these resources are publicly available over the Internet. They provide biomedical researchers with tools to facilitate different tasks, including, for instance, genetic sequence analysis and alignment, protein annotation or structural studies. As regards, there is a plethora of biomedical databases covering different areas including biomedical literature, genetic disorders, macromolecular structures or rare diseases. In this scenario, there is a need for novel tools to gather and organize all these resources. In most cases, the resources can be unknown to those researchers that may benefit from their use. This circumstance emphasizes the need for developing advanced tools to facilitate the localization and use of such resources.

In order to disseminate and provide a unified access to these resources, it is required to build an index of existing biomedical resources. Such an index should not only contain information regarding the location, inputs, outputs and service invocation procedures, but also a complete semantic description of their functionalities. Currently, there is not such an index readily available to the biomedical community.

resources owned by the INB. Valuable external resources from other institutions are not included in this index.

Other additional examples of resource indexes over the Internet are, for instance (i) the database collection mentioned by Galperin [1], and (ii) the list of resources available at [2]. The former reference [1] compiles a collection of publicly available molecular biology databases. This compilation is reviewed and published periodically every year. The latest version gathers a list of almost one thousand databases. The latter are classified into different categories according to their contents sorted in alphabetical order. For each database, the name, a short description of their contents and the link to the database is provided. The list of resources provided by (ii) is not only composed by databases but also tools, services and links to literature. More than two thousand resources have been classified into this list, depending on its functionality. In this paper, we present a novel semi-automatic method to create and maintain an index of biomedical resources. Figure 2 outlines the proposed method to create the Resourceome index.

2. Methods

A common problem for readily available indexes is that resources are annotated manually or they are not annotated at all. Creating or updating those indices by hand is a time-consuming task. Thus, it is a significant challenge to develop new methods and tools to facilitate the annotation and indexing of the resources. Another problem is the lack of standard domain models covering all required types of resources and tasks. All existing indexes use their own classification criteria or taxonomies that are neither complete nor do they use standard terminology to name their components.

Therefore, the first step would be the creation of a domain model or ontology not only covering the taxonomy of concepts but also all relevant relationships between the concepts. Domain knowledge provided by the relationships might contribute to a better understanding of the functionality provided by the resources and can be useful for resource classification and to enhance searches. According to Cannata et al. [3], ontologies are a suitable tool to annotate the resources and building the index. Firstly, they propose the need for creating an ontology including high-level concepts, attributes, and standard relationships between concepts. Next, biomedical researchers must be provided with a mechanism to maintain and extend the ontology with subconcepts to better describe their resources. Finally, the index built upon such ontology should be disseminated, enabling software engineers to create interfaces for searching and managing the resources.

First of all, it is necessary to obtain a list of the names of all the available resources that may be of interest for the biomedical community. Resources include databases, tools, and services. This list is created by an automated software agent that extracts the resource names from different online sources. Additional resource names can be added manually to the list by human annotators, if required. Once the list of resources has been obtained, it is necessary to extract detailed descriptions about their functionalities. This information can be usually found in abstracts of published papers or technical reports describing the resources. In many cases it is not necessary to examine the entire documents to extract this information, since abstracts often contain all the relevant information about the functionality provided by the resources. Once we obtain all the relevant abstracts, these are stored in a document repository. It is possible to store more

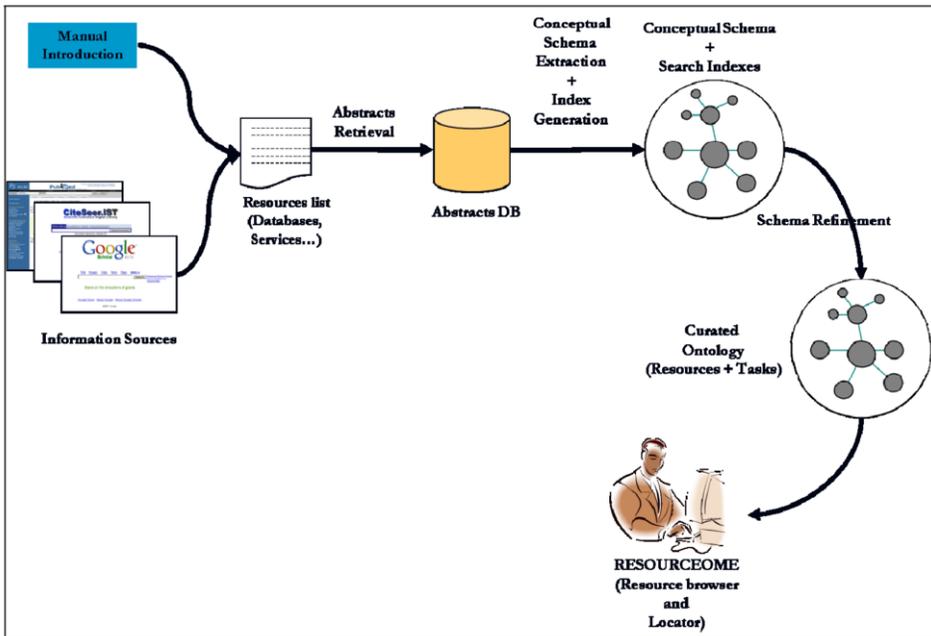


Figure 2. Overview of the “Resourceome” index creation method

than one abstract for a given resource in the database thus providing a more detailed description of the resource. Using the abstracts database, we automatically create a conceptual schema that describes the domain covered by the abstract database. This schema includes concepts – resources and tasks – as well as relationships between concepts. In this schema, relationships are of a single type – i.e. they denote which resources are useful for concrete tasks. The schema is extracted automatically from the selected abstracts using an original method developed by the authors and published elsewhere [4]. The method, based in different Artificial Intelligence (AI) and natural language techniques – including tokenizers, probabilistic part-of-speech taggers and transition networks – has been successfully used to bridge together schema-less, non-structured sources with relational databases. For further details on this issue see [4].

Our method works as follows. Each abstract is subdivided into phrases which are treated separately. Resources names, functionalities and all the relevant information is extracted by applying a special transition network built for this purpose. Besides the conceptual schema, it is necessary to create an index collection that relates the available resources to concepts in the conceptual schema. Using these indices, it is possible to browse the generated model to search for all the available tools to carry out a given task or databases containing a concrete type of data. Domain knowledge provided by the relationships can be exploited to enhance the searches.

Once the conceptual schema has been created, it is necessary to perform a refinement process to ensure the quality of the generated schema. The refinement task includes the removal of non-relevant concepts and relationships and naming concepts in the schema using a standardized terminology. To perform this task – which must be carried out by experts in the domain – it is possible to use widely accepted biomedical

vocabularies such as, for instance, the Unified Medical Language System (UMLS) [5], the Gene Ontology [6] or the Human Gene Nomenclature [7].

The refined model or ontology can be visualized by users using a graphical ontology viewer. We are currently developing an application programming interface (API) to provide access and search services for the Resourceome. Regarding the maintenance of the Resourceome, it is possible to incrementally update the index with new resources simply by adding new abstracts to the database. The system automatically creates the indices for the added resources. Next, the existing conceptual schema is automatically updated with concepts extracted from the new abstracts, not currently belonging to the conceptual model. Finally, experts can manually update the refined model – if required – to reflect the changes introduced by the addition of new resources to the Resourceome.

3. Results

In order to evaluate the model proposed in the previous section, we performed an experiment using a controlled input. We selected such kind of input since the AI and Natural Language Processing (NLP) techniques used in the development require some tuning to be adapted to the biomedical domain. This tuning is normally performed using controlled inputs to the system. We have designed this experiment targeted to tune and refine the model prior to execute more comprehensive tests with different data sources. The starting point was the list of resources offered by the EBI through its official website, previously mentioned. The list is organized into three categories (databases, tools and services). The names of fifty resources were randomly selected among the three categories and they were compiled into a candidate list. For each name of the list, the most relevant paper describing the resource was chosen from a public repository. For this experiment, we have used the ISI Web of Knowledge[®] [8] as the information source.

Once all the abstracts were manually retrieved, several AI and NLP techniques were applied to extract the relevant information contained in the abstracts. We have previously worked on a similar topic targeted to database integration and information retrieval [4]. For the present work, we have used the same approach for concept extraction and index generation. The target was focused in obtaining the name of the resources and their functionality. In the case of the databases, the detection of the functionality was substituted by the type of the data contained in the records. Additionally, extra information such as inputs, outputs, as well as relations with other resources was also searched. However, such information was only detected in few cases. Abstracts do not usually contain complete descriptions of the resources and such extra extraction is not always possible. Hence, we determined that an analysis of the full papers would be needed. Information extracted was automatically stored into a database specially built for this purpose in order to facilitate data analysis.

Some interesting results were discovered after analyzing the extracted data. First, the names of the fifty resources were properly extracted from the abstracts as well as their functionalities. A limited set of types or categories of resources has been established. In 94% of the cases, this category was correctly obtained from the functionality. Sometimes more than one category was found for a concrete resource. In such a case, both types were stored. These categories or concepts have been used to build a taxonomy for allowing the automatic classification of resources.

4. Conclusions and Future Research

The new model proposed is still in its early stages. Additional improvements and experimentation are needed. Results obtained in the preliminary experiments are promising and they suggest that the approach might be useful. Several information sources available on the Internet are currently being analyzed to be incorporated within the system in the next stages. Although some processes exposed before have been manually performed at this stage of the project, we are considering new approaches for automating the methods. We believe that the proposed method is an adequate starting point to organize the idea of a biomedical Resourceome.

Besides annotating and organizing the biomedical resources, the Resourceome might be a valuable tool to facilitate the construction of biomedical workflows. A workflow is a loopless forward chaining graph of available resources targeted to achieve a complex task. Workflows are usually created manually using tools such as Taverna [9], or Semantic Moby [10]. These tools provide workflow designers with a framework to create and execute scientific workflows in a graphical and intuitive manner. The main drawback of the manual workflow creation approach is that workflow designers need to have a good knowledge of the available resources as well as their required inputs and outputs. This approach can be suitable for simple or *ad-hoc* purposes, but it is not practical for more complex scenarios. The possible applications of creating automatically such workflows for the proposed Resourceome in different scientific areas is a great challenge for extending and applying the model presented.

Acknowledgments

The present work has been funded by the Ministry of Education, Spain (OntoMineBase project, reference TSI2006-13021-C02-01) and the E.C. (the INFOBIOMED NoE (FP6-2002-IST-507585) and the ACGT integrated project (FP6-2005-IST-026996)).

References

- [1] M.Y. Galperin, "The Molecular Biology Database Collection: 2007 update", Nucl. Acids Res. 35:D3-D4, 2006.
- [2] Bioinformatics Links Directory. http://bioinformatics.ca/links_directory/ (last accessed Feb. 15th, 2008).
- [3] N. Cannata, E. Merelli, R.B. Altman, "Time to organize the bioinformatics resourceome", PLoS Comput Biol 1(7):e76, 2005.
- [4] M. Garcia-Remesal, V. Maojo, J. Crespo, H. Billhardt, "Logical Schema Acquisition from Text-Based Sources for Structured Biomedical Sources Integration". AMIA Annu Symp Proc. 2007;259-263, 2007.
- [5] D.A.B. Lindberg, B.L. Humphreys, A.T. Mccray, "The Unified Medical Language System", Methods of Information in Medicine 32(4):281-291, 1993.
- [6] The Gene Ontology Consortium, "Gene Ontology: tool for the unification of biology", Nature Genetics 25(1):25-29, 2000.
- [7] S. Povey, R. Lovering, E. Bruford, M. Wright, M. Lush, H. Wain, "The HUGO Gene Nomenclature Committee (HGNC)", Human Genetics 109(6):678-680, 2001.
- [8] ISI Web of Knowledge. <http://isiknowledge.com> (last accessed Feb. 15th, 2008).
- [9] T. Oinn, M. Addis, J. Ferris, D. Marvin, M. Senger, M. Greenwood, et al, "Taverna: a tool for the composition and enactment of bioinformatics workflows", Bioinformatics 20(17): 3045-3054 Nov 22, 2004.
- [10] M.D. Wilkinson, D. Gessler, A. Farmer, L. Stein, "The BioMOBY Project Explores Open-Source, Simple, Extensible Protocols for Enabling Biological Database Interoperability". Proc Virt Conf Genom and Bioinf (3):17-27, 2003.

Improving Pain & Symptom Management for Advanced Cancer Patients with a Clinical Decision Support System

Krista ELVIDGE¹

Health Informatics Lab, Dalhousie University, Halifax, Canada

Abstract. In palliative medicine, healthcare providers aim to provide end-of-life cancer patients with a plan of care to minimize pain and manage symptoms, while providing psychosocial and educational support to patients and their families. Unfortunately, it has been reported that patients often experience unnecessary suffering due to ineffective symptom management as they near end-of-life. Recent advances in health informatics have motivated healthcare institutions to take advantage of clinical decision support systems that assist healthcare providers with evidence-based decision making for pain and symptom management. In this paper, we present a unique clinical decision support system that incorporates case-based reasoning and evidence-based standards of care. It is anticipated that this user-friendly, web-based CBR system will improve decision making for pain and symptom management for end-of-life cancer patients.

Keywords. Care pathways, case-based reasoning, clinical decision support system, end-of-life cancer care, symptom management guidelines

1. Introduction

Cancer is the leading cause of morbidity and premature death in Canada. With a growing and aging population, we can expect the incidence rate of cancer to steadily increase over time, with 38% of women and 44% of men developing cancer during their lifetime [1,2].

Many patients who develop advanced cancer will experience its accompanying symptomatology that precedes death, including physical and psychological suffering [3], and the array of symptoms they contend with will significantly impact their ability to carry out activities of daily living, overall well being, and quality of life [5].

In Canada, palliative care is an integral component of cancer care which can improve the quality of life of patients and their families through the prevention of suffering, symptom assessment, and delivery of effective treatment [3].

In this paper, we present a clinical decision support system (CDSS) to assist healthcare providers with decision-making to improve pain and symptom management in end-of-life cancer care.

¹ Corresponding Author: K. Elvidge. Dalhousie University, Halifax, Canada. kelvidge@dal.ca

1.1. Symptoms Experienced by Patients with Advanced Cancer

Advanced-cancer patients frequently experience an array of physical and psychosocial symptoms when they near end-of-life; the most common clinical symptoms include pain, dyspnea, nausea, fatigue, anorexia, cachexia, anxiety, and depression [3]. To measure the effectiveness of pain and symptom management, healthcare providers have begun to adopt reliable, validated assessment tools such as the Edmonton Symptom Assessment Scale (ESAS) and the Palliative Performance Scale (PPS).

2. Problem Description

Although assessment tools have been widely adopted in palliative care, 30% of Canadian patients with advanced cancer receive suboptimal, ineffective symptom management; and as a result they continue to suffer unrelieved pain and other cancer-related symptoms as they near end-of-life [1]. According to many studies, effective pain and symptom control can be achieved in 90% of patients with advanced cancer using existing knowledge and resources. Despite the World Health Organization's widely accepted goal for palliative care, substandard symptom management remains a ubiquitous public health problem which significantly diminishes quality of life for patients with advanced cancer [1].

Recent studies in palliative care have extensively explored barriers to adequate symptom management for patients with advanced cancer, and have found that the majority of these barriers implicate deficits in knowledge and attitude of healthcare providers as main obstacles to effective symptom management. In Canada, the four main barriers impeding excellent symptom management for advanced-cancer patients include: (a) inadequate education and practical training in symptom management and end-of-life care for healthcare providers, (b) inadequate standards of care (CPGs and clinical pathways) and lack of evidence-based practice, (c) fragmentation of palliative care services and limited integration of healthcare providers across the continuum of palliative care, and (d) failure to reliably capture a record of the quality of care delivered to patients with advanced cancer [1-3].

3. Methodology

Clinical practice guidelines and integrated care pathways have been successfully implemented to promote adherence to an accepted standard of care, while encouraging clinical judgement and freedom when deviation from the standard of care is deemed necessary [2]. Several studies have proposed that clinical practice guidelines and integrated care pathways could be similarly applied to palliative care to alleviate the common barriers which currently prevent the delivery of effective pain and symptom management for end-of-life cancer patients [2,3]. Furthermore, healthcare providers have recommended that the barriers which impede the delivery of effective pain and symptom management could be improved when healthcare providers adopt intelligent clinical decision support systems to assist with the complex task of clinical decision making for end-of-life cancer care [6].

3.1. Clinical Decision Making in End-of-Life Care

In the knowledge intensive field of palliative care, a vast amount of information can be derived from clinical experience and archived in a repository for future usage. This collection of information is a priceless asset which can assist healthcare providers with clinical decision making in palliative care; however, managing this volume of information can be overwhelming, and certainly memorizing all the information is virtually impossible. To take advantage of this voluminous information, healthcare providers have begun to implement intelligent clinical decision support systems to support clinical decision making [6].

3.2. Clinical Decision Support Systems

A clinical decision support system presents healthcare providers with an interactive, computerized system that uses data and models to generate information to support clinical decision making. The three main components of a decision support system are a knowledge base, an inference engine, and a user interface. A knowledge base typically contains domain expertise which may be represented as clinical guidelines, decisional rules, and records of past patient cases. An inference engine is a computer program which processes information using systematic inference steps, similar to the decisional steps employed in the human thought process, and uses one or more reasoning methodologies [7].

Case-based reasoning is one such reasoning methodology that is proving to be especially valuable in mitigating error in diagnosis and patient management in palliative medicine. More recent advancements in health informatics, especially in the artificial intelligence (AI) techniques of data mining and case-based reasoning, present an opportunity to improve pain and symptom management through the implementation of case-based reasoning (CBR) medical decision support systems [7].

3.3. Research Objective

Our aim is to develop a CDSS to support clinical decision-making for the delivery of effective pain and symptom management for end-of-life cancer patients. The system will employ case-based reasoning methodology, and serve as a medium from which healthcare providers can leverage knowledge elicited from previously-solved patient cases to assist with the development of effective symptom management care plans for end-of-life cancer care.

4. Case-Based Reasoning System Architecture

The design of the presented case-based reasoning system was motivated by Aamodt and Plaza's (1994) [8] CBR cycle, whereupon the case-based reasoner must perform four sequential processes: case retrieval, case reuse, case revision, and case retention. To enable these processes, there are four main tasks that must be employed: (1) case representation, (2) case indexing and storage, (3) case retrieval, and (4) case adaptation and learning. Below, we provide the design details of how these tasks were accomplished in our CBR system.

4.1. Case Representation

In this clinical decision support system, CBR was used to support clinical decision making for pain and symptom management in cancer care. Patient cases are represented by several feature attributes that describe a problem (e.g., mild pain) and the solution to the problem (e.g., 200-800 mg ibuprofen). The feature attributes that formed the problem description were: gender, age, cancer type, palliative performance scale (PPS), palliative performance scale-level, and the symptoms that were experienced by advanced cancer patients while in the Victoria Hospice Society palliative care unit.

For this research, simulated case solutions were derived using collaborative care plans and evidence-based symptom management guidelines. These case solutions were vetted by physicians in palliative care services, Capital District Health Authority. The solution attributes (care intervention categories) which formed case solutions were: patient assessment, pharmacological considerations, pharmacological therapy, non-pharmacological therapy, patient activity, psychosocial considerations, and patient/family education.

4.2. Case Indexing and Storage

In our system, patient cases (N = 276) were stored in a MySQL database and MySQL indices were created to speed-up database searching. The primary key of the database was the patient ID number to ensure that every patient record was unique. Indices were also created on the attribute columns to speed-up case comparison and retrieval.

4.3. Case Retrieval

4.3.1. Case Retrieval using the Nearest Neighbor Algorithm

In this CBR system, healthcare providers enter the feature attributes of a new patient case on an electronic web form, then perform a search query to retrieve the four most similar patient cases that are stored in the case library. The resultant set of cases is presented to the user in order of descending similarity. During case retrieval, similarity computation is calculated by applying the nearest-neighbor algorithm, Eq. (1), to find the most similar cases from the case library.

$$\text{Similarity}(T,S) = \sum f(T_i, S_i) \times w_i \quad (1)$$

In Eq. (1), w_i is the weighting factor of each individual feature attribute (i), S_i represents stored cases of each individual attribute, and T_i is the target case attributes. Weighting factors (w_i), or importance values, were determined for each feature attribute using Spearman's rank correlations and discriminant analysis. Each case has a set of feature weights which is added to produce a weighted sum. The nearest-neighbor algorithm uses a distance calculation to compare the Euclidian distance between two feature vectors (weighted sum of the new query case with the weighted sum of cases stored in the case library) to determine how similar two cases are by comparing their features [9]. Euclidean distances have been determined using confusion matrices (Figure 1).

Pain/Dyspnea ESAS	NONE	MILD	MODERATE	SEVERE
NONE	1	0.75	0.5	0.25
MILD	0.75	1	0.75	0.50
MODERATE	0.50	0.75	1	0.75
SEVERE	0.25	0.50	0.75	1

Figure 1. Confusion matrix for pain and dyspnea

4.3.2. Case Retrieval using Inductive Decision Tree/Nearest Neighbor Algorithm

In the presented system, we developed a second retrieval algorithm whereby we combined an inductive decision tree with a nearest-neighbor algorithm to optimize the efficiency of the CBR system. The inductive decision tree is used to retrieve a selection of indexed cases for subsequent similarity matching by the nearest-neighbor algorithm, which applies a similarity metric to return ranked cases in descending order of similarity to the target case [9].

The ID3 algorithm is used to induce a decision tree by ranking attributes according to their importance in classifying the data. This algorithm ranks attributes using an entropy measure, and is first applied to the complete corpus of cases to determine the attribute with the highest entropy. This attribute becomes the top-most decision point of the tree and is referred to as the root node. Subsequently, this process is applied recursively down the tree until all cases have been classified and indexed in leaves or until there are no further attributes to incorporate [9].

4.4. Case Adaptation and Learning

Our presented CBR system currently does not incorporate case adaptation; however, we are exploring the method of compositional adaptation, which is predicated upon combining the most salient solution components from multiple past cases to derive a final composite solution which is more representative of a user's target case query. In medical case-based reasoning, case adaptation continues to be the main challenge to the development of a clinical decision support system which applies the complete CBR method. In the present system, end-users can store a new case into the case library (case learning) using an electronic web form. Upon submission, this information is stored as a new case in the case library, and so it would be available for future case queries.

5. Evaluating System Efficacy and Clinical Validity of Case Solutions

In an external evaluation process, key informants from palliative care examined our simulated case solutions for clinical appropriateness, deficiencies, erroneous care protocols, and whether case solutions represented local best practice and clinical practice guidelines. Future investigation will incorporate an RCT to evaluate the clinical validity of our CDSS in several clinical environments (PCU, ICU, hospice, home palliative care).

6. Concluding Remarks

The Canadian Cancer Society anticipates that death due to cancer will soon surpass cardiovascular disease, to become the leading cause of death in Canada. With this trend, there will be an increasing demand for palliative care services to provide cancer patients with effective relief from pain and other debilitating symptoms that are often experienced near end-of-life. According to WHO, substandard symptom management remains a ubiquitous public health problem, despite the fact that effective symptom management could be achieved for approximately 90% of advanced-cancer patients using existing knowledge such as standards of care.

In this paper, we presented a unique CDSS which incorporates evidence-based standards of care, and takes advantage of case-based reasoning to support clinical decision-making in end-of-life cancer care. Currently, our prototype CDSS is still under development; however, we believe that full implementation of our CBR system could improve clinical decision-making in end-of-life cancer care, ultimately improving symptom management and quality-of-life for end-of-life cancer patients.

Our research warrants further investigation to explore the technique of compositional adaptation, and the feasibility of implementing a clinical decision support system using case-based reasoning to support end-of-life cancer care in palliative care services, Capital District Health Authority, Nova Scotia, Canada.

Acknowledgements

We would like to acknowledge our partnership with the Victoria Hospice Society and the University of Victoria, Canada. We extend thanks to Michael Downing, Jack Shaw, and Francis Lau, for sharing the palliative database, and to Dr. Paul McIntyre and Dr. Robert Horton, who provided valuable insight during the clinical validation process.

References

- [1] Cancer Care Nova Scotia. Cancer management guidelines. <http://www.cancercare.ns.ca> 2007 January.
- [2] Dudgeon D, et al. Managing continuity through collaborative care plans: A study of palliative care patients. <http://www.chrsf.ca> 2007 June.
- [3] Foley K, Gelband H. *Improving Palliative Care for Cancer*. Washington, D.C.: National Academy Press, 2001.
- [4] Bruera E, Pereira J. Recent developments in palliative cancer care. *Acta Oncologica* 1998; 37(7): 749-757.
- [5] Babka O, Whar S. Case-based reasoning and decision support systems. Paper presented at the IEEE International Conference on Intelligent Processing Systems; 1997 October; Beijing, China.
- [6] Fortier P, Michel H, Sarangarajan B, Dluhy N, O'Neill E. A computerized decision support aid for critical care novice nursing. *Proceedings of the 38th Hawaii International Conference on System Sciences*; 2005; Hawaii.
- [7] Aamodt A, Plaza E. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications* 1994; 7(1): 39-59.
- [8] Althoff K, et al. Case-based reasoning for medical decision tasks: The INRECA approach. *Artificial Intelligence in Medicine* 1998; 12(1): 25-41.
- [9] Craw S, Rowe R. Case-based reasoning for tablet formulation: A novel technology to re-use past formulations. <http://www.comp.rgu.ac.uk/staff/smc/papers/pharmaventures-article.pdf> 2007 July.

Collaboration Patterns in an Online Community of Practice in Oral Medicine

Göran FALKMAN^{a,1}, Marie GUSTAFSSON^a,
Olof TORGERSSON^b and Mats JONTELL^c

^a *School of Humanities and Informatics, University of Skövde,
P.O. Box 408, SE-541 28 Skövde, Sweden*

^b *Department of Computer Science and Engineering, Göteborg University*

^c *Institute of Odontology, Sahlgrenska Academy, Göteborg University*

Abstract. SOMWeb is an online collaboration system based on Semantic Web technologies, which is used for knowledge sharing and dissemination within an oral medicine community in Sweden. Based on a previous study of the use of SOMWeb, general patterns of interaction and communicative activities involved in community collaboration have been identified. The patterns for one such activity, distance consultation, are described and modeled using techniques from the Pragmatic Web. It is also shown how patterns could inform system design.

Keywords. Collaboration, Information management-dissemination, Knowledge-based systems, Modeling, Dentistry, Human Factors

1. Introduction

Evidence-based medicine (EBM) entails integrating the expertise of the individual clinician with the best medical evidence obtainable from different knowledge sources [1]. Central to EBM is the externalization of clinical practice knowledge of the individual clinician into diffused knowledge (e.g., protocols and clinical guidelines), together with the possibility to exploit explicit knowledge sources [2]. Although social communication is essential for both the externalization and exploitation of knowledge, interaction and communicative actions are often neglected in HIS design [3]. One way to promote the externalization process is to provide IT-support for the communities of practice (CoP) [4] that is formed by practitioners of a medical domain, taking advantage of the practitioners' passion for their profession, their ambition of learning how to do it better, and their mutual interest of advancing the level of knowledge within their domain.

A prerequisite of IT-supported externalization of clinical knowledge is the representation of knowledge in a computer-processable manner. Within the area of knowledge representation, there has been much focus on ontologies during the past ten years [5]. However, despite much effort, the adoption of ontologies within the medical domain has turned out to be more problematic and slower than many had hoped [6]. Open

¹ Corresponding Author: Göran Falkman, School of Humanities and Informatics, University of Skövde, P.O. Box 408, SE-541 28 Skövde, Sweden; E-mail: goran.falkman@his.se.

problems include how a group agrees on how a concept should be represented and how an ontology of the individual can be reconciled with an ontology of the community. The Pragmatic Web [7] [8] is a recent extension to the Semantic Web² in which these problems are addressed through the use of reusable patterns of meaning and knowledge flows, patterns in which emphasis is put on the context in which knowledge occurs [9].

SOMWeb is an online CoP for the Swedish Oral Medicine Network (SOMNet) [10]. SOMWeb is based on the Resource Description Framework (RDF) and the Web Ontology Language (OWL) of the Semantic Web, and uses ontologies for representing examination templates, values used in filling in templates, and for representing community data (information about users, meetings, cases, and news). In our previous research, we have studied the clinicians' use of SOMWeb, as the first step in the elicitation of contextual factors and communicative activities involved in knowledge sharing processes in oral medicine [11]. The study suggested improvements to SOMWeb based on taking contextual elements related to the users, their activity, and their environment into consideration. We have also studied the possibility of using ideas from the Pragmatic Web to identify pragmatic patterns within SOMNet [12]. It was shown how activities in SOMNet could be seen as instances of the community activation pattern in [13]. In this paper, we describe how the pattern for a general activity within SOMNet can be modeled and how a specific instance of this pattern, the pattern for distance consultation, can be put into use in SOMWeb.

2. Methods and Materials

2.1. Study Context

SOMNet functions as a community of practice within oral medicine in Sweden. As such, SOMNet promotes the collection of cases for subsequent analysis and harmonization within the field. Through regularly held teleconference meetings, SOMNet provides means for distance consultations and learning for a broader audience. In this, SOMNet is an important platform in the interaction between oral medicine and related medical specialties. As of November 2007, SOMNet has about 70 members in Sweden.

SOMWeb is an online CoP for communication and knowledge sharing and dissemination within SOMNet. Members can add cases, with associated pictures, that are discussed at SOMNet meetings. The activities supported by SOMWeb include browsing cases, meetings, and members of the system; looking at presentations of individual cases and meetings (past, current, or future); adding and administering the cases one owns; administering meetings; reading news; and using the discussion forum, with threads both associated with individual cases and topics that are more general.

2.2. Methods

Since 2004, researchers from computer science and interaction design have regularly observed the SOMNet teleconference meetings. An online questionnaire was also provided to the members of SOMNet. Users were polled for, e.g., perceived usefulness of the system for performing certain tasks and for reasons having/not having used certain functionality. A selection of SOMWeb users has also been interviewed.

² <http://www.w3.org/2001/sw/>

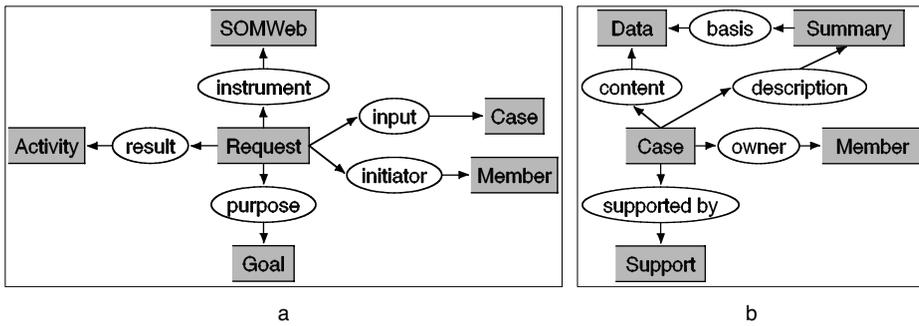


Figure 1. The activity request pattern (a) and case pattern (b).

3. Results

3.1. Collaboration Patterns in SOMWeb

Based on the study of the collaboration within SOMNET and the use of SOMWeb, we have identified a number of *collaboration patterns* that are involved in the SOMWeb activities that are the result of a user requesting some input on a specific case.

In general, such an activity is initiated by a member using the SOMWeb portal to issue a request, with the purpose of fulfilling some goal and with a specific case as the starting point. Figure 1 (a) depicts this *activity request pattern* using a conceptual graph notation [14]. Our study of the use of SOMWeb informs us that the goal of a member is seeking consultation on a case, wanting to share information about a case, or creating a discussion about a case.³ The case in question is part of the community resources, in that it has previously been added to the SOMWeb database. Every case has a member as its owner. The contents of a case consist of data, which in this case means examination data and photos, and a description of the case is provided in the form of a summary that is constructed using the template-based natural language generation component of SOMWeb. The case may also be supported by other resources, e.g., similar cases or various types of evidence. This general *case pattern* is depicted in Figure 1 (b).

The result of an activity request is an instance of the *activity pattern* (Figure 2). As

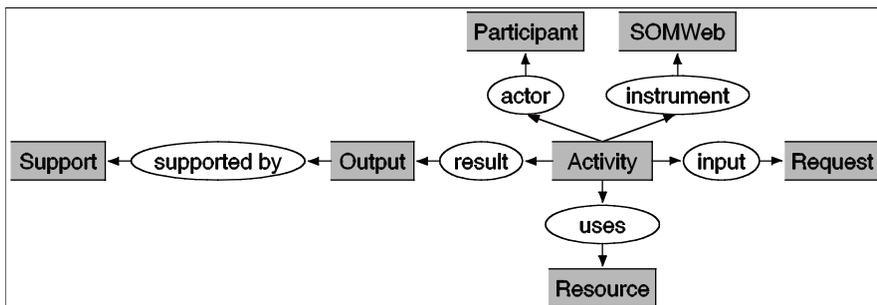


Figure 2. The activity pattern describes the implementation of an activity request.

³ For the reason of brevity, we will not give an explicit presentation of the conceptual graph type hierarchy used in the presented patterns, e.g., formally defining 'creating a discussion' to be a subclass of 'goal'.

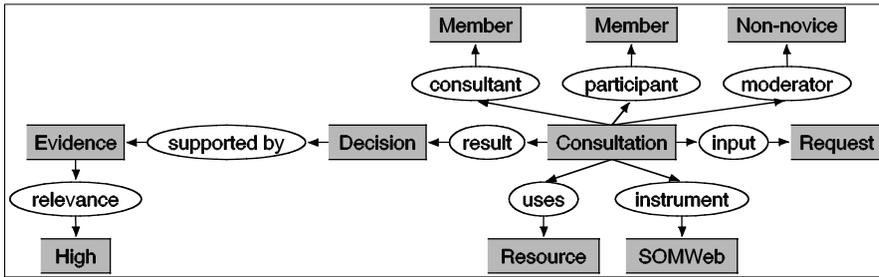


Figure 3. The consultation pattern is a specialization of the activity pattern.

was noted earlier, the resulting activity is case consultation, case sharing, or case discussion. Common to all three types of activities are that they take an activity request as input (including the purpose, initiator, and input case of the request), some action is performed by the participants of the activity, during which resources of the community (members and stored cases) are used, and that they result in some output, which is supported by some kind of support. In all this, SOMWeb is instrumental.

A specialization of the activity pattern is the *consultation pattern*, as depicted in Figure 3. The consultation pattern stipulates three types of participants: one moderator of the consultation, which must not be a novice user, one consultant, and any number of other participants, all of which are members of the community. The output of a consultation activity is a decision. In practice, there are three types of decisions: establishing a diagnosis, deciding on a treatment plan, and general advice on how to proceed with a case (including deciding on a follow-up meeting). The pattern also requires that the decision should be supported by evidence of high relevancy (and quality).

3.1.1. Patterns in Use

We describe the use of the collaboration patterns using the following plausible scenario, thereby also giving some hints on how the patterns could inform system design.

Consider a member, M, an expert in certain lesions of the oral mucosa that has added cases to SOMWeb several times before. When M wishes to add a new case for consultation to the system, M is presented with an input form, which is generated from an OWL description of the content of a case. Since the case will be the basis for a decision regarding diagnosis or treatment, the system could adopt a relatively strict policy regarding M entering fundamental case data. After completion of the form, the system could initiate a matching process to look for previous cases added by M on the same or similar subjects, and for identifying members within the same area of expertise. These cases and members become parts of the case's supporting resources.

When the case is added to the system, M has the option of assigning it to one of the upcoming teleconference meetings. However, the consultation pattern states that there should be a moderator of the consultation, and, by default, this will be the chairperson of the proposed meeting. The designated chairperson is automatically notified by email of M's consultation request. The chairperson could then decide not to schedule the consultation on the proposed meeting, e.g., because very few of the members that ought to be interested in this case can attend the meeting. When the consultation finally has been assigned to a meeting, invitations could be sent out to interested parties. Before the meeting takes place, a discussion thread can also automatically be set up in one of the discussion forums provided by SOMWeb, with the consultant as the moderator.

During the meeting, the system can use the consultation pattern to help guiding the chairperson/moderator through the discussions, towards a decision that fulfills the goal of M's request. The system could also act as an advisor during the consultation activity, by providing reminders to the supporting cases and cuing participants that have been identified to be especially relevant for the case to address the meeting.

At the end of the meeting, the system could secure that a decision is being made and that the outcome is recorded and communicated to the community. This also entails the search for relevant external medical evidence and the assessment of its quality.

4. Discussion

Patterns supporting interaction and communicative actions in clinical knowledge processes can be one part of promoting EBM. Patterns put emphasis on the caregivers' interests and needs and on the contexts and processes in which different evidence occurs, thereby giving the clinicians the experience that EBM services are beneficial. Thus, pattern-based services have the potential of being integrated into clinical practice and therefore offer good possibilities for long-term use and increased patient benefit. In addition, general reusable patterns of clinical knowledge and decision-making could be turned into improved health care strategies, which over time provide new evidence, adding to the foundation of EBM.

In [13], generic collaboration patterns are used to describe community memory activation. In terms of our consultation pattern, the *goal pattern* would be to establish either a diagnosis or a treatment plan, supported by the best medical evidence available; *communication patterns* would include getting approval of the submitted case from the designated chairperson; *task patterns* would include recording the outcome of the meeting; an *information pattern* would be how to search for relevant medical evidence and assess its quality; and a *meta-pattern* would describe what to do if no relevant evidence supporting the decision can be found.

While we find the approach of [13] to be a good starting point, it does not directly address the questions of the origin of patterns and on what level of abstraction patterns should be defined. One way forward could be to connect the elements of community collaboration with an ontology over organizational elements. For example, in identifying and describing the goal patterns, the normative structure of the relationships that exist among members of an organization (e.g., values, norms, and role expectations) [2] could be useful. Further, the behavioral structure of an organization can be described in terms of activities and interactions between members, and, thus, could be used for identifying relevant communication patterns. Finally, in identifying and describing communication and information patterns, the technologies (material resources as well as technical knowledge and skills of members) available are a factor to consider.

A SOMNet meeting is an example of a multidisciplinary medical team meeting (MDTM), where the team members meet to review patient cases, establish a diagnosis, and decide on the most appropriate treatment plan for the patient [15]. The processes associated with a MDTM system are: (1) pre-meeting activities; (2) case presentation; (3) case discussion, including negotiation and reinterpretation of findings; (4) deciding on the diagnosis and treatment; (5) recording of the outcome; and (6) post-meeting activities. The same processes can be discerned in a SOMNet meeting. The processes 2–5 of a MDTM could serve as the appropriate level of abstraction (or granularity) for identifying and defining sub-patterns within the consultation part of the consultation pattern.

5. Conclusions

We have described how a pattern for a general collaboration activity within a CoP in oral medicine can be modeled, and how a specific instance of this pattern, the pattern for distance consultation, could be put into use in SOMWeb, an existing online CoP. The case study indicates that research on collaboration patterns within the Pragmatic Web can be useful in this, but that it would benefit from the addition of research on organizational ontology and multidisciplinary medical team meetings.

The long-term objectives of our research are to identify, represent, and make use of recurring patterns of interaction and communicative actions in clinical knowledge processes within oral medicine, patterns that could inform the design of online CoP. In the short term, we intend to continue studying the use of the SOMWeb system. As for patterns, the next step is to refine the consultation pattern and formalize the conceptual graph representation of patterns by extending the SOMWeb ontologies.

Acknowledgements

This work is funded by the Swedish Governmental Agency for Innovation Systems (VINNOVA), research grant 2006-02792.

References

- [1] W. Rosenberg and A. Donald, Evidence based medicine: An approach to clinical problem solving, *Brit. Med. J.* **310** (1995), 1122–1126.
- [2] M. Stefanelli, The socio-organizational age in artificial intelligence in medicine, *Artif. Intell. Med.* **23** (2001), 25–47.
- [3] E. Byrne and J. Gregory, Co-constructing local meanings for child health indicators in community-based information systems: The Uthukela district child survival project in KwaZulu-Natal, *Int. J. Med. Inform.* **76** (2007), 78–88.
- [4] E. Wenger, *Communities of practice: Learning, meaning, and identity*, Cambridge University Press, Cambridge, U.K. and New York, N.Y., 1998.
- [5] T. R. Gruber, A translation approach to portable ontologies, *Knowl. Acquis.* **5** (1993), 199–220.
- [6] A. Yu, Methods in biomedical ontology, *J. Biomed. Inform.* **39** (2006), 252–266.
- [7] M. P. Singh, The Pragmatic Web, *IEEE Internet Comput.* **6** (2002), 4–5.
- [8] M. Schoop, A. de Moor and J. L. Dietz, The Pragmatic Web: A manifesto, *Commun. ACM* **49** (2006), 75–76.
- [9] A. de Moor, Patterns for the Pragmatic Web, in Proc. 13th International Conference on Conceptual Structures, ICCS 2005, *Lect. Notes Artif. Intell.* **3596** (2005), 1–18.
- [10] M. Gustafsson, G. Falkman, F. Lindahl and O. Torgersson, Enabling an online community for sharing oral medicine cases using Semantic Web technologies, in Proc. 5th International Semantic Web Conference, ICSW 2006, *Lect. Notes Comput. Sci.* **4273** (2006), 820–832.
- [11] M. Gustafsson and G. Falkman, Modeling Contexts of Knowledge Sharing in an Online Community for Oral Medicine, in Proc. Doctorial consortium, *CONTEXT 2007*, 2007, pp. 30–43.
- [12] G. Falkman, M. Gustafsson, O. Torgersson and M. Jontell, Towards Pragmatic Patterns for Clinical Knowledge Management, in: Proc. 2nd International Conference on the Pragmatic Web (ICPW'07), Tilburg, The Netherlands, October 22–23, 2007, pp. 65–74.
- [13] A. de Moor, Community memory activation with collaboration patterns, in Proc. 3rd International Community Informatics Conference, CIRN 2006, 2006, pp. 1–18.
- [14] S. Polovina and J. Heaton, An Introduction to Conceptual Graphs, *AI Expert* **7** (1992), 36–43.
- [15] B. Kane and S. Luz, Multidisciplinary medical team meetings: An analysis of collaborative working with special attention to timing and teleconferencing, *Comput. Support. Coop. Work* **15** (2006), 501–535.

An Automated Personalised Intervention Algorithm for Remote Patient Monitoring

Joanna FURSSE^{a,1}, Malcolm CLARKE^a, Russell JONES^b, Sneha KHEMKA^c,

Genevieve FINDLAY^c

^a*Brunel University, UK*

^b*Chorleywood Health Centre, UK*

^c*BUPA, UK*

Abstract. An automated personalised intervention algorithm was developed to determine when and if patients with chronic disease in a remote monitoring programme required intervention for management of their condition. The effectiveness of the algorithm has so far been evaluated on 29 patients. It was found to be particularly effective in monitoring newly diagnosed patients, patients requiring a change in medication as well as highlighting those that were not conforming to their medication. Our approach indicates that RPM used with the intervention algorithm and a clinical protocol can be effective in a primary care setting for targeting those patients that would most benefit from monitoring.

Keywords: Remote patient monitoring, telemedicine, e-Health, telehealth.

Introduction

It is estimated that there are over 17 ½ million people living with a long term condition in the UK accounting for over 80% of GP consultations. With an aging society these numbers are only set to increase further [1]. Evidence has shown that even relatively small reductions in blood pressure and blood glucose in hypertensives and diabetics reduce their clinical risk significantly [2, 3]. It is vital that healthcare services are able to provide these gains for patients at an affordable cost. Remote Patient Monitoring (RPM) has been identified as a potential tool to manage this demand for health care [4]. To date there have been many projects that have evaluated the technology [5], however

¹ Corresponding Author: Joanna Fursse, Brunel University, Uxbridge, Middlesex, UB8 3PH, UK; E-mail: malcolm.clarke@brunel.ac.uk

most have concentrated on detecting and managing the acute exacerbations that arise from chronic disease (particularly CHF) rather than determine effective ways to manage long term the condition itself. This study aims to develop and evaluate methods that exploit both RPM in the patients' homes and protocol-based clinical interventions to achieve sustained improvement in disease measurements for three long-term conditions: chronic heart failure (CHF); Type 2 diabetes mellitus; and essential hypertension.

1. Methods

Our study was based in Chorleywood Health Centre, a medium sized general practice to the NW of London. From a total population of 6000 registered patients, 724 were determined to have CHF; type II diabetes mellitus; and essential hypertension, some having two or more. 173 were chosen randomly and invited to take part in the study. Of these 51 accepted the invitation to participate in the study.

Each participant was provided with an RPM unit for a 12 week period, the study having three such cohorts over a nine month period. We had a further set of 3 patients having the equipment for the entire nine month period and a set of 3 patients with no equipment. We used a commercial RPM system that consisted of a unit with a touch screen device to allow manual entry of certain data (glucose) and attached peripheral devices to measure automatically appropriate medical values. Those with CHF were given weighing scales, SpO₂, and BP unit; those with Diabetes a glucometer; and those with Hypertension a BP unit. Each participant was asked to enter daily physiological measurements. Targets were set for each group based on best practice guidelines [6, 7, 8] : BP of 140/85mmHG for non diabetic patients and 130/80mmHG for patients with diabetes, HbA1c of 7% (8.6 mmol/L). CHF was managed by monitoring weight, BP and SpO₂ for significant change.

The data was sent via the participant's telephone line to a server held at the Health Centre and could be accessed and viewed via a website. A chart of the data was created for each patient and this was hyperlinked into the EPR (electronic patient record). The primary healthcare team reviewed each patient whose data lay beyond the threshold defined by our intervention algorithm at regular case conferences and clinical interventions were agreed.

2. Intervention Protocol

We developed a sophisticated personalised automated intervention algorithm within the project to improve the accuracy with which patients requiring intervention were recognised compared to existing systems based on a simple threshold. The specific goals were to reduce the number of false positive indications and to adapt automatically when a clinical intervention was made so that no further indications were given until sufficient time had elapsed for the intervention to take effect.

Our approach was to smooth the raw data by using a median filter, which is effective at removing outlier values. We then wished to apply an adaptive threshold that was based on the patient's data, responded to an intervention and would eventually provide an indication should the patient data remain above the long term target value.

Such an adaptive threshold can be achieved by applying an exponential curve having an initial value that is 2 standard deviations above the current value, and

decaying over a 14 day period to the long term target value. The threshold is seen to model well changes in the data when the patient responded to an intervention.

The algorithm is shown applied to a hypertensive patient in figure 1 where the final target is set at 140/85. The smooth curves indicate the boundaries within which the data should lie to achieve best clinical outcome as described in the Map of Medicine [9]. To avoid needless triggers and alarms the boundaries are adjusted upward at the beginning of monitoring and after the introduction of an intervention, the value being set to 2 SD above the current value.

A case conference run with the doctors and nurses of the team reviews the collected data and the clinical notes. Data is observed on a daily basis but unless swings are extreme, data is reported on a weekly basis and reviewed for possible intervention every two weeks. Figure 1 shows the successful addition of a calcium blocker to better control the systolic blood pressure.

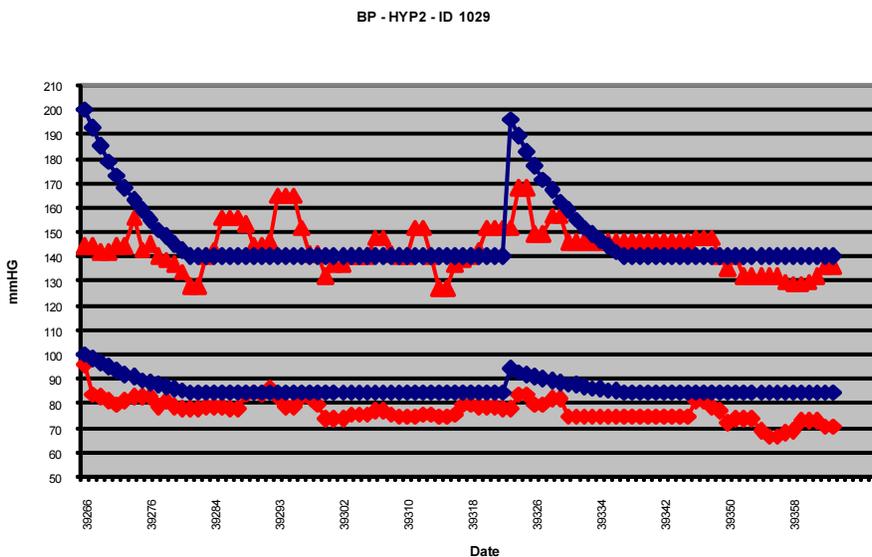


Figure 1. Personalised automated intervention protocol for a hypertensive patient

Figure 2 shows a diabetic patient unwilling to accept her newly diagnosed diabetes and treatment until there is eventual control of the blood sugar following repeated intervention.

Figure 3 shows the daily blood glucose readings for a patient known to have long term management issues of diabetes. On each clinical intervention, the threshold is reset to a value higher than the current value and it can be seen to reduce back to the target value over a 14 day period. The patient refused nursing advice to rotate his injection site, his persistence with one site being due to his difficulty in coping when in his wheel chair – he is disabled because of a diabetic neuropathy – and his faith in his private diabetologist. Observing his results changed his thinking. The final change was to introduce Metformin.

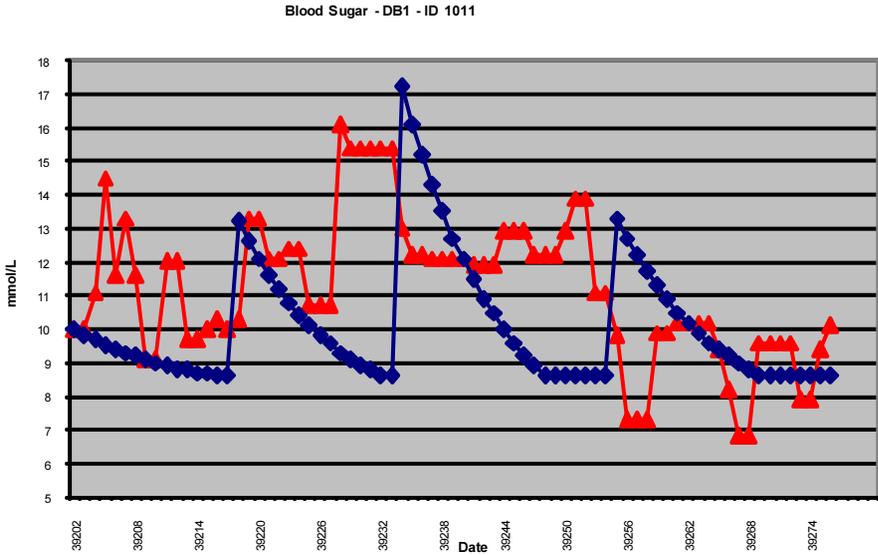


Figure 2. Personalised automated intervention protocol for diabetic patient

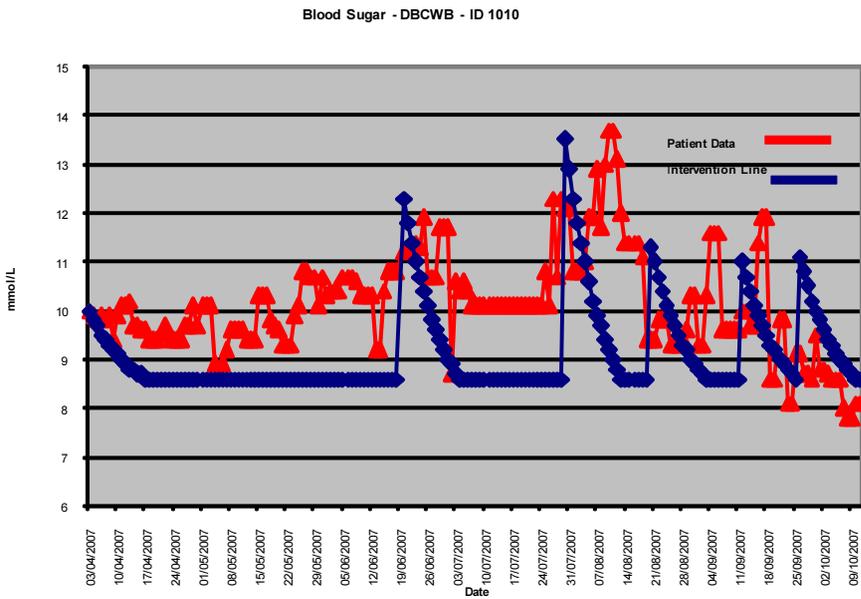


Figure 3. Personalised automated intervention protocol for diabetic patient

3. Preliminary Results

After 6 months of the study, 2 cohorts have each completed the 12 week trial, giving data on 29 patients (mean age 70, 17 female). 11 patients had essential hypertension as

their primary long term condition, 8 had chronic heart failure, 7 had type II diabetes mellitus, 2 had chronic heart failure and type II diabetes mellitus, and 1 had all three conditions. 9 participants withdrew from the study before they were due to be given the RPM units because of clinical or social reasons and further participants were recruited to replace. None withdrew during monitoring.

The algorithm has prompted clinical intervention in 11 patients (37% of the 1st and 2nd cohort): 64% of patients with CHF, 18% of the Hypertensive group and 18% of the diabetes group. The average time elapsed before first intervention was 47 days (SD 21). Primarily these interventions (72%) resulted in changes to medication and health advice, however one hypertensive patient was referred for a pace maker after the discovery of a bradycardia due to heart block, and one CHF patient had two emergency hospital admissions during the 12 week period. These interventions have resulted in a mean shift reduction of 12mmHG Systolic and 2mmHG Diastolic in the hypertensive group.

4. Discussion

The personalised threshold algorithm was developed to be based on the patient's data, clinical target values and clinical interventions so that it might easily be applied automatically within a RPM system. When linked to their primary healthcare team, the need for clinical intervention is easily recognised and thus provides better and earlier disease management. Limiting the amount of time required to review the data by both reducing the number of alerts and filtering out those patients not requiring intervention contains work load and makes adoption of the system more likely.

Results suggest that 4 weeks is sufficient time in which to recognise a need to intervene clinically and that in 12 weeks it is possible to intervene enough to affect a change toward target. This reduces both the disruption of patients' lives and the cost of the service because the equipment can be reused quickly.

In the UK, the GMS contract has made general practice the key organisation to care for chronic disease. But the clinical question remains – on whom and when should we intervene? A single occasional measure of a physiological parameter might fail to reveal poor care and deterioration in a patient's illness. A clinical response to a single reading of blood pressure can lead to excessive dosages and harm. In contrast in this study, clinical interventions were agreed upon by the team and carried out by both nurses and doctors in their clinics and in the patients' homes. The value of RPM was established quickly in the minds of the team as the results were reviewed and made any additional clinical work acceptable. The answering of clinical questions that arose often needed blood tests to be done; ECGs recorded; ambulatory ECG and blood pressure recordings made; and referrals arranged. The combination of technology and clinical skills possible in general practice can do such work effectively and it is doubtful that other services such as the polyclinic can be close enough to the patient to make a difference.

5. Conclusions

RPM used with automated personalised intervention algorithms and a clinical protocol was found to be very effective in the long term management of patients with chronic disease in this primary care setting. It was particularly effective in the case of newly diagnosed patients, patients needing change in medication, those unwilling to comply,

and those with poor long term control, as it supported an aggressive targeted strategy for intervention. By improving the intervention algorithm, fewer false positive patients were detected, giving the users confidence to act on the information.

By shifting the focus of care away from managing acute exacerbations, clinicians can use RPM to recognise those of their patients who would most benefit from monitoring. The clinical information provided by RPM allows both healthcare professional and patient to confront issues together and resolve difficulties. Awareness of the physiological measurements and learning their value from the clinical team empowers patients in gaining self-determination in understanding and managing their own illness resulting in benefits for themselves, their healthcare professionals, and the health economy.

Competing interests: The remote monitoring devices have been in part funded by BUPA, although BUPA has no financial interest in the project's outcomes.

References

- [1] DOH, Improving Chronic Disease Management, 2004, Available at: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4_075214
- [2] Prospective Studies Collaboration, Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies, *The Lancet* 2002; **360**: 1903-1913
- [3] Stratton IM, Adler AI, Andrew H, Neil W, Mathews DR, Manley SE, Cull CA, Hadden D, Turner RC, Holman RR. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study, *British Medical Journal* 2000; 321: 405-412
- [4] Barlow J, Singh D, Bayer S, Curry R. A systematic review of the benefits of home telecare for frail elderly people and those with long-term conditions, *Journal of Telemedicine and Telecare* 2007;13:172-179
- [5] Bensink M, Hailey D, Wootton R. A systematic review of successes and failures in home telehealth: preliminary results, *Journal of Telemedicine and Telecare* 2006; 12 (Suppl. 3): S3:8-16
- [6] Williams B, Poulter NR, Brown MJ, Davis M, McInnes GT, Potter JF, Sever PS, MCG Thom S. Guidelines for management of hypertension: report of the fourth working party of the British Hypertension Society 2004 – BHS IV, *Journal of Human Hypertension*, 2004; 18: 139-185
- [7] McIntosh A, Hutchinson A, Home PD, Brown F, Bruce A, Damerell A, Davis R, Field R, Frost G, Marshall S, Roddick J, Tesfave S, Withers H, Suckling R, Smith S, Griffin S, Kaltenthaler E, Peters J, Feder G. Clinical Guidelines and evidence review for Type 2 diabetes: management of blood glucose, Sheffield: SchARR, University of Sheffield, 2001
- [8] McIntosh A, Hutchinson A, Home PD, Brown F, Bruce A, Damerell A, Davis R, Field R, Frost G, Marshall S, Roddick J, Tesfave S, Withers H, Suckling R, Smith S, Griffin S, Kaltenthaler E, Peters J, Feder G. Clinical Guidelines and evidence review for Type 2 diabetes: management of blood glucose, Sheffield: SchARR, University of Sheffield, 2001
- [9] Map of Medicine. Available at <http://www.mapofmedicine.com/>

Syntactical Negation Detection in Clinical Practice Guidelines

Stefan GINDL ^{a,1}, Katharina KAISER ^a and Silvia MIKSCH ^{a,b}

^a *Institute of Software Technology and Interactive Systems, Vienna University of Technology, Vienna, Austria*

^b *Department of Information and Knowledge Engineering, Danube University Krems, Krems, Austria*

Abstract. In clinical practice guidelines (CPGs) the medical information is stored in a narrative way. A large part of this information occurs in a negated form. The detection of negation in CPGs is an important task since it helps medical personnel to identify not occurring symptoms and diseases as well as treatment actions that should not be accomplished. We developed algorithms capable of Negation Detection in this kind of medical documents. According to our results, we are convinced that the involvement of syntactical methods can improve Negation Detection, not only in medical writings but also in arbitrary narrative texts.

Keywords. Negation detection, clinical practice guidelines, natural language processing

Introduction

Negation is an important part of inter-human communication. It can be used to invert concepts and to show refusal of opinions. The concept of negation is a universal concept in all languages and very important in the medical field. Detecting negations in natural language is a difficult task, but in the medical scope it is easier: Medical language is much more restricted than narrative speech [1]; a physician will not use stylistic elements such as double negation extensively to write reports or patients histories.

In the medical scope Negation Detection is currently only applied to very simple texts (e.g., radiology reports). In our work, we primarily focus on the more complex text type of clinical practice guidelines (CPGs). These are “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances” [2]. In CPGs negation is crucial not only for facts that do not apply (e.g., patient has no pain), but also for actions that should not be accomplished (e.g., do not take this drug). In contrast to simpler texts, we need

¹ Corresponding Author: Stefan Gindl, Institute of Software Technology and Interactive Systems, Vienna University of Technology, Favoritenstrasse 9-11/188, A-1040 Vienna, Austria; E-mail: stefan.gindl@gmx.at

algorithms dealing with the syntax (e.g., tenses in active and passive voice, parts-of-speech) of the English language.

In the following section we give an overview over existing methods of Negation Detection. In the main part of the work we describe and evaluate an approach of Negation Detection using syntactical methods tailored to the special characteristics of CPGs.

1. Related Work

Besides general work on negation in natural language (e.g., [3]) we will discuss relevant work addressing Negation Detection in the scope of medical language.

NegEx [4] is a simple algorithm for detecting negated findings and diseases in radiology reports. Negation triggers are classified in triggers with preceding negated concepts and those with succeeding concepts. After a replacement of the concepts by UMLS terms, the negated ones are detected.

In the work of Mutalik et al. [5] UMLS concepts are also identified in a first step. Then, a lexical scanner using regular expressions is applied for trigger detection and classification in preceding and succeeding triggers. With this information a parser provides the original concepts with the negation information leading to the output of the NegFinder algorithm.

Elkin et al. [6] use an algorithm with a rule base to decide which medical concepts are negated in clinical documents. Here, stop words (e.g., “other than”) determine the scope of a negation trigger.

Patrick et al. [7] use SNOMED CT to identify negated concepts. Thereby, pre-coordinated phrases (e.g., “no headache”, SNOMED CT concept id 162298006) and concepts explicitly asserted as negative by negation phrases are identified. To identify the latter rule-based algorithms similar to [4] and [6] were implemented.

Huang and Lowe [8] recently developed a hybrid approach combining regular expression matching with grammatical parsing to detect negations in clinical radiology reports.

Aronow and Feng [9] have developed a method for Negation Detection to be applied for document classification. Thereby, they determine the scope of negation triggers by conjunctive phrases. All phrases connected by such conjunctions are regarded as negated phrases.

CPGs differ from medical reports or discharge summaries, which are used by the algorithms presented. In CPGs the language is not as restricted as it is in these other documents. They are more like prosaic writings, which complicate the development of simple algorithms. Still, they are not as complicated as free text since sophisticated stylistic elements such as the double negation are not used. In the following section we explain our approach, called *NegHunter*.

2. NegHunter – A Method to Detect Negated Concepts

Our strategy when developing our method, called *NegHunter*, was to classify negations in CPGs according to identified negation types. The reason for this is that negations within CPGs are strongly varying from each other. It is not easily possible to keep the number of negation triggers within manageable limits, thus a syntactical approach was

used. This means that grammatical elements of the English language are used to decide whether a phrase is negated or not. For this purpose, the tenses in both active and passive voice as well as parts-of-speech are used.

The starting point of the *NegHunter* algorithm is the detection of negation triggers. Whereas other algorithms use a relatively high number of different negation triggers, *NegHunter* gets along with a rather small number of triggers. The reason for this is the way *NegHunter* handles the different negation types. We have selected a number of universal triggers and classified their behaviour in narrative texts.

2.1. Negation Classes

We have come up with five negation classes according to our study of CPGs in the literature: (1) adverbial negation, (2) intra-phrase triggered negation, (3) prepositional negation, (4) adjective negation, and (5) verb negation. In the following, we will discuss these five classes in detail.

1. The Adverbial Negation

This is the most frequent type of negation. Triggered by “not” and “never”, negated concepts appear in combination with a verb. Via the tense of the verb we decide, whether the sentence is written in active or passive voice. With this information, we interpret the three preceding or succeeding noun phrases as negated. We use the number of three noun phrases as we receive the best results with it. The following two sentences show examples of sentences in active and passive voice, both with triggers and their negated phrases.²

“Guideline developers do not recommend chemotherapy.” (active voice)

“Chemotherapy is not recommended.” (passive voice)

2. Intra-Phrase Triggered Negation

These are negations in which the trigger is included in the noun phrase. “No” and “without” act as triggers. The following sentence shows an example:

“Evidence obtained from at least one well-designed study without randomisation.”

3. Prepositional Negation

In this negation type triggers are followed by prepositional phrases, often introduced by the prepositions “of” or “from”. The phrases following the preposition are considered as negated. Here, we have the three noun triggers “lack”, “absence”, and “freedom” as well as the adjective trigger “free”. The result of the detection process could be as follows:

“Patients with good performance status, ... , and the absence of systemic disease.”

4. Adjective Negation

This type of negation uses adjectives as negation triggers. We have identified, for example, the term “ineffective” and can interpret the first noun phrase before the trigger as the relevant phrase. For example:

“Recommendation indicates at least fair evidence that the service is ineffective or that harm outweighs benefit.”

5. Verb Negation

Some verbs are also negation triggers themselves. We identified the verbs “deny”, “decline”, and “lack”. The following sentence shows the effect of such a trigger:

“Information on final patient outcomes was also lacking.”

² Negation triggers are underlined; solid frames signalize negated phrases.

2.2. Assigning prepositional phrases

In some cases, not the entire negated information gets tagged with the algorithms described above. For instance, prepositional phrases (which are by themselves not negated) appearing after a negated phrase need to be handled apart. We proceed with this problem by tagging all prepositional phrases that follow a negated phrase. This ensures that no information concerning the negation is lost. The following sentence shows an output result with two prepositional phrases following an intra-phrase triggered negation:

“Requires availability of well conducted clinical studies but no randomised clinical trials on the topic of recommendation.”³

3. Evaluation

For evaluation purposes we used a Java-implementation of our algorithms. To receive the syntactical information from the guideline documents necessary for our algorithms we used the MetaMap Transfer (MMTx) program. MetaMap is “a program [...] to map biomedical text to the [UMLS] Metathesaurus or, equivalently, to discover concepts referred to in text” [10]. MMTx makes this program available for researchers in an adaptive way. Besides the concept assignment it also provides us with the syntactical information such as part-of-speech. We implemented our algorithm as a Java library that can also be used by and incorporated in other programs and applications. In the following we describe the evaluation process.

3.1. Training and Test Sets

We used a set of 18 CPGs from the medical speciality oncology for our development. Out of these 18 practice guidelines, we used four guidelines as training set for the analysis of occurring negations. By means of these documents we classified the negations and developed our algorithms. We used the remaining 14 CPGs for the evaluation.

3.2. Generation of Gold Standard Documents

We manually rated the sentences of all oncological CGPs to establish a “gold standard” against which the computerized algorithms could be compared. We processed 558 sentences containing 615 negated concepts and tagged both negation triggers and negated concepts. At the first glance, it may be irritating to have such a little more number of negated phrases than sentences containing negations. This is because there are many sentences containing a trigger, which does not aim a phrase in the same sentence, (e.g., “None available.”). We do not provide detection across sentence borders because the result is unpredictable and are not tackled in our methods conceptually.

³ Dotted frames signalize prepositional information.

3.3. Evaluation Techniques and Measures

For our evaluation, we processed the 14 guidelines with *NegHunter*. Afterwards, a hand reading was carried out to detect errors. We classified in true positives (TP), false positives (FP), false negatives (FN), and partially correct (PC) taggings, whereas the latter scored only 50 %.

Table 1. Evaluation results in the form of recall and precision measures of each negation type and their percentage of all occurring true positives (TP), false positives (FP), and false negative (FN) phrases.

Negation Type	Triggers (%)	TP (%)	FP (%)	FN (%)	PC (%)	Recall (%)	Precision (%)
Adverbial Negation	66.12	65.25	94.26	91	71.43	76.41	58.97
Intra-Phrase Tr. Negation	23.45	26.85	0	7	16.33	95.1	100
Prepositional Negation	7.66	3.36	0.82	0	4.08	100	89.47
Adjective Negation	0.65	1.18	0	0	0	100	100
Verb Negation	2.12	3.36	4.92	2	8.16	89.47	58.62
Overall	100	100	100	100	100	83.51	67.49

To qualify our measurement we used the statistical parameters of recall and precision. The recall measures the number of the correctly found phrases against all relevant phrases according to the gold standard. The value of precision measures the ratio of the number of correctly detected phrases to the number of all found phrases of the system. Table 1 shows a detailed listing of the performance of our implementation.

3.4. Analysis of Evaluation Results

NegHunter shows its strength in the handling of the intra-phrase triggered negation, the prepositional negation and the adjective negation. This is caused by the simple structure of these negations. In the case of the intra-phrase triggered and the prepositional negation, the negated phrase usually follows immediately after the trigger so it is nearly impossible to fail it.

The behaviour of the adverbial negation as well as the verb negation is much more complex. Here, it is possible that a phrase related with a trigger occurs at the diametrically opposite end of the sentence. In such a case, it is very difficult to identify this phrase, as *NegHunter* uses the range of three preceding or succeeding possible phrases for detecting the negated concepts.

Another problem is generated by MMTx itself. In some cases the part-of-speech is incorrectly assigned and this consecutively causes errors. For example, in the sentence

“... and interpreting studies that were not otherwise covered in existing syntheses or guidelines.”

MMTx recognises the noun phrase “studies” as a verb phrase, whereas “interpreting”, a verb phrase, is recognised as noun. This circumstance leads to a false tagging and the creation of both a FP and a FN.

4. Conclusion

With our presented algorithms, negated information occurring in CPGs can be detected on syntactical level using grammatical information of the English language such as tenses and parts-of-speech. This forms a basis for subsequent processing also on a semantic level. Further processing on a semantic level will be absolutely necessary, as, for instance, a negation trigger and a concept representing a symptom or disease may not imply the absence of this symptom or disease. Compare also the example of [4]:

“We did not treat the infection.”

“We did not detect an infection.”

where the first sentence does not indicate the absence of an infection, but the absence of treating it. Anyhow, using *NegHunter* can support an automated structuring of the information in order to, for instance, decide which therapies or drug regimens are best applied in patients with certain diseases and which are not recommended. This helps to sort out the treatment options and supports the medical personnel as well as patients in their decision-making.

Additionally, *NegHunter*'s negation classification allows users to augment the trigger set by themselves. Therefore, new triggers need to be assigned a negation class. *NegHunter* applies its rule base to these new triggers. This makes *NegHunter* portable to be applied on other document types as well as extensible and maintainable.

Acknowledgement

This work is supported by “Fonds zur Förderung der wissenschaftlichen Forschung FWF” (Austrian Science Fund), grant L290-N04.

References

- [1] Naomi Sager, Carol Friedman, and Margaret S. Lyman. *Medical Language Processing: Computer Management of Narrative Data*. Addison-Wesley Longman Publishing, 1987.
- [2] Marilyn J. Field and Kathleen N. Lohr. *Clinical Practice Guidelines: Directions for a New Program*. National Academies Press, Institute of Medicine, Washington DC, 1990.
- [3] Laurence R. Hom. *A Natural History of Negation*. University of Chicago Press, Chicago, Illinois, June 15, 1989.
- [4] Wendy W. Chapman, Will Bridewell, Paul Hanbury, Gregory F. Cooper and Bruce G. Buchanan. A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries. *Journal of Biomedical Informatics* 34:301-310 (2001).
- [5] Pradeep G. Mutalik, Aniruddha Deshpande and Prakash M. Nadkarni. Use of General-purpose Negation Detection to Augment Concept Indexing of Medical Documents: A Quantitative Study using the UMLS. *Journal of the American Medical Informatics Association (JAMIA)*, 8(6):598-609, 2001.
- [6] Peter L. Elkin, Steven H. Brown, Brent A. Bauer, Casey S. Husser, William Carruth, Larry R. Bergstrom and Dietlind L. Wahner-Roedler. A controlled trial of automated classification of negation from clinical notes. *BMC Medical Informatics and Decision Making*, 5(13), 2005.
- [7] Jon Patrick, Yefeng Wang and Peter Budd. Automatic Mapping Clinical Notes to Medical Terminologies. Proceedings of the 2006 Australasian Language Technology Workshop (ALTW 2006), pages 75-82, 2006.
- [8] Yang Huang and Henry J. Lowe. A novel hybrid approach to automated negation detection in clinical radiology reports. *J Am Med Inform Assoc*, 14:304–311, 2007.
- [9] David B. Aronow and Fangfang Feng. Ad-Hoc Classification of Electronic Clinical Documents. *D-Lib Magazine*, January 1997.
- [10] Alan R. Aronson. Effective Mapping of Biomedical Text to the UMLS Metathesaurus: The MetaMap Program. Proc. of the AMIA Symposium, pages 17-21, 2001.

Subjective usability of the CARDSS guideline-based decision support system

Rick GOUD^a, Monique W.M. JASPERS^a, Arie HASMAN^a, Niels PEEK^a
^a *Department of Medical Informatics, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands*

Abstract. Clinical decision support systems (CDSSs) differ from other health information systems in their aim to directly influence the decision-making behaviour of healthcare professionals. As a result, CDSSs face additional challenges with respect to user acceptance. The objective of this study was to investigate subjective usability of a guideline-based CDSS for outpatient cardiac rehabilitation. The system, named CARDSS, was previously found to be effective in improving guideline adherence of rehabilitation professionals in a cluster randomized trial. To assess CARDSS' usability, a modified version of the IBM Computer System Usability Questionnaire was sent to all 68 professionals from the 28 outpatient clinics that participated in the trial. The questionnaire was returned by 63 respondents (93%) from 27 clinics. Factors that influenced CARDSS' usability were identified using linear regression analysis. Analysis showed that professionals who managed to smoothly integrate the system with their daily routine were more satisfied with ease of system use. Furthermore, a positive attitude of respondents towards CDSSs in general and a better agreement with the content of the national guidelines were positively correlated to satisfaction with CARDSS' overall usability and each of its sub-domains.

Keywords. Decision support, Evidence based guidelines, Assessment-Evaluation

Introduction

Although clinical guidelines describe evidence-based best practices for specific healthcare conditions, adherence to guidelines in clinical practice is often found to be low [1]. Guideline-based computerized decision support systems (CDSSs) have proven to potentially be effective instruments for improving guideline adherence [2]. These CDSSs provide patient-specific recommendations to care professionals based on the guideline's recommendations. Despite of their potential benefit, however, CDSSs are still not widely adopted in healthcare [3].

Poorly designed CDSSs can lead to usability problems, users' dissatisfaction and may disrupt normal flow of clinical activities. So whether an information system will be adopted in medical practice depends, among other things, on the 'fit' between technology, user, and the tasks that need to be performed [4;5]. Evaluating users' satisfaction with the usability of the system in question is important to understand if and how such fit was achieved [6], also for CDSSs [7]. However, it is unknown what influences users' satisfaction with the usability of CDSSs [2] as CDSSs have been

predominantly evaluated using quantitative methods that can not explain why the system was or wasn't adopted [2;3].

Guideline-based CDSSs differ from other health information systems in that they aim to increase guideline adherence by influencing the behaviour of professionals. This makes that the factors influencing users' satisfaction with CDSS usability may also differ. Users' satisfaction with CDSS usability is presumably dependent on the general attitude of professionals towards being advised by a software system in their daily practice [5]. Novice professionals, who have little experience with prevailing protocols and guidelines, might be more open to advice than their more experienced colleagues. Also care professionals' attitude towards the guidelines underlying the CDSS may be critical to their satisfaction with a CDSS' usability. CDSSs that are well integrated into the existing working procedures and decision-making processes of healthcare professionals have proven to be more effective in terms of adherence to standards [8]. One may conjecture that the underlying mechanism is usability.

In this paper we report on a usability study that was carried out with a guideline-based CDSS for cardiac rehabilitation in the Netherlands, called CARDSS (Cardiac Rehabilitation Decision Support System). Cardiac rehabilitation is a multidisciplinary therapy that is provided after cardiac events (e.g. myocardial infarctions) and cardiac interventions (e.g. heart surgery). CARDSS actively guides professionals (primarily nurses and other paramedics) with conducting the needs assessment procedure for cardiac rehabilitation described in the national guidelines. This proceeds through a structured dialogue, prompting the professional to enter the necessary patient information and proposing the rehabilitation therapies that are considered appropriate for the patient in question ('consulting model', [9]). Following the recommendations in the literature, CARDSS also provides various additional patient information management services and CARDSS' design takes into account the working procedures specific to multidisciplinary outpatient care [8;10].

Recently a cluster randomised trial was carried out in 34 Dutch outpatient clinics to evaluate the effect of CARDSS on guideline adherence. The participating clinics either worked with an intervention version of the system that included decision support functionality, or with a control version that lacked decision support but otherwise had the same functionalities. Initial results of this trial show that the CDSS increased adherence to the guideline; detailed results will be published elsewhere.

To gain understanding in why and in what way CARDSS was effective, we evaluated which factors specific to CDSSs influenced our users' satisfaction with CARDSS. This paper describes the results of this evaluation, which can help to understand which factors are important for the adoption of CDSSs in multidisciplinary outpatient settings aiming to influence the behaviour of non-physicians.

1. Methods

To evaluate the usability of CARDSS, a questionnaire was developed based on the IBM Computer System Usability Questionnaire (CSUQ) [11]. The CSUQ is developed to measure satisfaction of software users with respect to system usability, and contains 19 items related to users' satisfaction with ease of system use, the quality and clarity of information and error messaging provided by the system, and the system's interface quality. Each of the items is rated on a 7-point Likert scale. We excluded two items concerning the error handling of CARDSS in the calculation of scores for the

'information quality' subscale as we considered them non-specific to users' satisfaction with a CDSS' information quality.

To identify factors that influenced CARDSS' usability, questions were added referring to the respondents' clinical experience with cardiac rehabilitation, the way they integrated CARDSS in their working procedures, and the time increase per patient caused by using the system. In addition, questions were added to assess the respondents' agreement with the content of the national cardiac rehabilitation guidelines (12 items) and their general attitude towards the use of CDSSs in healthcare (2 items). These latter items were also rated on a 7-point Likert scale. Also basic questions on the respondents' age, gender, and computer literacy were included. For details of the entire questionnaire, we refer the reader to [12].

Questionnaires were sent to all 28 outpatient clinics that completed the trial with CARDSS and were asked to let professionals that worked with CARDSS fill in and return the questionnaire. The questionnaire took about 15 minutes to complete. Respondents were allowed to answer 'not applicable' if they felt they had insufficiently worked with CARDSS to answer the question concerned.

For each respondent we calculated the average scores for overall usability (19 items), ease of system use (subscale, 8 items), information quality (subscale, 5 items) and interface quality (subscale, 3 items). Average scores were also calculated for professionals' agreement with the content of the national guidelines and their attitude towards CDSS in general. To analyze which factors influenced CDSS subjective usability, we performed univariate linear regression analysis adjusting for age, gender, study arm (intervention or control), and computer literacy of the respondents to correct for possible confounding. To correct for multiple testing a p-value of 0.01 was considered statistically significant.

Table 1. Summary of questionnaire results

	<i>n/a</i>	Value
Age, years (mean \pm SD)	–	42.8 \pm 8.2
Male gender (%)	–	32%
Self-judged computer experience	–	
Low literacy (%)		11 (18%)
Moderate literacy (%)		34 (54%)
High literacy (%)		18 (29%)
Clinical experience, years (median [interquartile range])	–	5 [2.5-14]
Integration of CARDSS into clinical workflow	3	
Concurrent: use CARDSS during patient visit (%)		16 (27%)
Serial: fill in paper during patient visit and use CARDSS afterwards (%)		44 (69%)
Time spend on needs assessment procedure per patient, minutes (median)	5	40-50
Time increase caused by CARDSS, minutes (median [interquartile range])	9	10 [0-10]
Agreement with content of cardiac rehabilitation guidelines (mean \pm SD)	3	5.07 \pm 0.72
Attitude towards use of CDSSs in general (mean \pm SD)	3	5.65 \pm 1.26
Overall usability (mean \pm SD)	–	5.10 \pm 0.85
Ease of system use (mean \pm SD)	–	5.14 \pm 1.08
Information quality (mean \pm SD)	2	5.05 \pm 0.87
Interface quality (mean \pm SD)	3	5.10 \pm 1.18

2. Results

63 out of 68 (93%) cardiac rehabilitation professionals from 27 (out of 28) outpatient clinics returned the questionnaire (18 intervention centres, 41 respondents; 9 control centres, 22 respondents). Table 1 presents the summary results for the questionnaires.

For the agreement with the content of cardiac rehabilitation guidelines and the general attitude towards use of CDSSs a score (near) 1 indicates respectively a low agreement and negative attitude, while a score (near) 7 indicates respectively a high agreement and positive attitude. For the satisfaction with overall usability and each of its subscales a score (near) 1 indicates a low satisfaction and a score (near) 7 indicates a high satisfaction. There were no significant differences in characteristics and scores between respondents that worked with the intervention version and those who worked with the control version of CARDSS.

Table 2 shows the results of the univariate linear regression analyses on factors influencing user satisfaction with CARDSS' overall usability and its subscales.

Table 2. Results of univariate linear regression analyses to identify factors that influence usability

	Ease of system use (B ± SE)	Inform. quality (B ± SE)	Interface quality (B ± SE)	Overall (B ± SE)
Age, per 10 years increase	-0.19 ± 0.17	-0.12 ± 0.14	-0.13 ± 0.19	-0.15 ± 0.13
Female sex	0.05 ± 0.30	0.14 ± 0.24	0.32 ± 0.32	0.10 ± 0.23
Intervention arm	0.09 ± 0.29	0.22 ± 0.23	0.31 ± 0.32	0.17 ± 0.23
<i>Computer literacy</i>				
Low literacy ‡	--	--	--	--
Moderate literacy	0.09 ± 0.38	0.34 ± 0.30	-0.06 ± 0.43	0.09 ± 0.30
High literacy	0.29 ± 0.42	0.37 ± 0.33	-0.22 ± 0.47	0.19 ± 0.33
<i>Clinical experience †</i>				
4 years or less	0.78 ± 0.45	0.40 ± 0.37	0.28 ± 0.55	0.62 ± 0.36
5 to 12 years	0.40 ± 0.42	0.54 ± 0.35	0.23 ± 0.51	0.47 ± 0.33
more than 12 years ‡	--	--	--	--
<i>Time increase by CDSS †</i>				
No time increase	1.22 ± 0.39 *	0.25 ± 0.38	1.18 ± 0.54	0.83 ± 0.31 *
1 to 10 minutes	0.81 ± 0.36	0.44 ± 0.36	1.08 ± 0.51	0.75 ± 0.29
more than 10 minutes ‡	--	--	--	--
<i>Workflow Integration CARDSS †</i>				
Concurrent use	1.08 ± 0.39 *	0.57 ± 0.34	0.75 ± 0.47	0.71 ± 0.31
Serial use ‡	--	--	--	--
Attitude decision support †	0.48 ± 0.09 *	0.27 ± 0.08 *	0.56 ± 0.11 *	0.39 ± 0.07 *
Agreement guideline content †	0.73 ± 0.18 *	0.69 ± 0.13 *	0.97 ± 0.19 *	0.75 ± 0.12 *

†: Adjusted for Age, Sex, Intervention arm, and Computer literacy

‡: reference category

*: Statistically significant at level $p \leq 0.01$

3. Discussion

In this study we analyzed which factors were related to subjective usability of CARDSS, a CDSS for outpatient cardiac rehabilitation. Several statistically significant associations were found.

Rehabilitation professionals that use the CARDSS system in their clinic are free to consult the system simultaneously with patient visits or to work on paper during visits and consult the system afterwards. Simultaneous consultation will often be more efficient as it avoids duplicate registration (on paper and electronically). Moreover, when the system is consulted afterwards, it may ask for patient items that were not discussed during the visit, creating an awkward situation. It is not surprising, then, that simultaneous users found the system more easy to use, as this is known to be related to integration of a system with users' workflow patterns [13]. Similarly, professionals who reported that using the system caused no time increase were more satisfied with CARDSS' ease of use. In this case, however, it is unclear whether poor workflow

integration caused extra time and therefore dissatisfaction with the system, usability problems led to poor integration and time increase, or perhaps a mixture of both scenarios occurred.

We also found that CARDSS users who held a positive attitude towards CDSSs in general, or better agreed with the content of the national guidelines, found the system more usable than others. These results touch upon the concepts of usability and usefulness. High system usability does not directly imply that users likewise would perceive a system as useful. In the present case, however, subjective usability and subjective usefulness appear to be firmly related. It should be noted though that our usability study was conducted after users had worked with the system for at least six months, and we do not know their earlier opinions. So, our results could suggest that users who generally perceived CDSSs as useful better appreciated the usability of CARDSS, but also that users who were satisfied with CARDSS developed a positive attitude towards CDSS in general. Similarly, agreement with the guidelines underlying CARDSS may have caused users to find the system more usable, but it may also have been the other way round.

In this study we also analyzed the relation between clinical experience and subjective CDSS usability as we hypothesized that novice professionals, who have little experience with prevailing protocols and guidelines, might be more open to advice than their more experienced colleagues. Although our analyses indeed show that the clinical experience of professionals is negatively correlated to satisfaction with CARDSS' usability, this correlation was not found to be statistically significant.

A limitation of this study is that questionnaires were only filled in by users from clinics that completed the CARDSS trial, and not by users from clinics that dropped out of the trial. At the start of the trial, four outpatient clinics that were assigned to the control arm of our study quit their participation because they reported that the benefits of using CARDSS without receiving guideline-recommended therapies did not outweigh the effort of learning to work with CARDSS. These users worked with CARDSS too short a time to judge its usability. However, as their reasons for not using the system were related to experimental conditions of usage during the trial and not to properties of the system itself, we do not believe that their exclusion has biased our findings.

The statistical analyses of this study assume that the points on a discrete Likert scale are equidistant, and that averaged responses from such a scale approximately follow a Normal (Gaussian) distribution. The latter assumption may be easily violated at the extremes of the response interval (1 through 7). Although most responses clearly indicate positive agreement (attitude/satisfaction), extreme responses were rare, and therefore these assumptions seem warranted in the present investigation.

In our study, some factors showed no statistically significant relation to the individual domains of subjective usability. However, the statistical power of our analysis is limited by our 63 respondents. Sample size calculation shows that we could have detected a difference between groups in subjective usability of 0.83 for CARDSS' workflow integration, and 1.01 for computer literacy, clinical experience, and time increase by CDSS, when allowing Type I and Type II error risks of 0.01 and 0.20, respectively, and using the average standard deviation of 0.97. Therefore, if these factors do have an effect on subjective usability, the actual differences between groups are below one. We consider this limitation acceptable as such a difference would in many cases be smaller than one Lickert-scale unit.

4. Conclusion

Several systematic reviews have tried to identify factors that are related to the effectiveness of CDSSs in terms of improving adherence to standards [2;8]. To be effective in this sense, a system must first be adopted by practitioners, but there is little knowledge on factors that influence CDSS adoption [3]. This study contributes to the understanding of such factors.

The results of this study show that the subjective usability of an active, guideline-based CDSS consulted by non-physicians in an outpatient setting is related to workflow integration, general attitude towards CDSSs, and agreement with the content of the guidelines that underlie the knowledge base.

Based on our findings we recommend CDSS developers and implementers to pay special attention to CDSS workflow integration to increase chances on successful system adoption. To further improve CDSS adoption, we recommend CDSS implementers to develop and apply educational strategies to improve sceptical future users' general attitude towards CDSSs and agreement with the guideline underlying the CDSS. These recommendations apply to active, guideline-based CDSSs consulted by non-physicians at outpatient clinics, but they may well generalize to other settings.

References

- [1] Grimshaw J, Eccles M, Tetroe J. Implementing clinical guidelines: current evidence and future implications. *J Contin Educ Health Prof* 2004;24 Suppl 1:S31-S37.
- [2] Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA* 2005;293(10):1223-38.
- [3] Kaplan B. Evaluating informatics applications--clinical decision support systems literature review. *Int J Med Inform* 2001;64(1):15-37.
- [4] Ammenwerth E, Iller C, Mahler C. IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study. *BMC Med Inform Decis Mak* 2006;6:3.
- [5] Trivedi MH, Kern JK, Marcee A, Grannemann B, Kleiber B, Bettinger T, et al. Development and implementation of computerized clinical guidelines: barriers and solutions. *Methods Inf Med* 2002;41(5):435-42.
- [6] Beuscart-Zephir MC, Brender J, Beuscart R, Menager-Depriester I. Cognitive evaluation: how to assess the usability of information technology in healthcare. *Comput Methods Programs Biomed* 1997 Sep;54(1-2):19-28.
- [7] Gardner RM, Lundsgaarde HP. Evaluation of user acceptance of a clinical expert system. *J Am Med Inform Assoc* 1994 Nov;1(6):428-38.
- [8] Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ* 2005 Apr 2;330(7494):765.
- [9] Musen MA, Shahar Y, Shortliffe EH. *Clinical Decision-Support Systems*. In: Shortliffe EH, Cimino JJ, editors. *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*. 3 ed. New York: Springer; 2006.
- [10] Shiffman RN, Liaw Y, Brandt CA, Corb GJ. Computer-based guideline implementation systems: a systematic review of functionality and effectiveness. *J Am Med Inform Assoc* 1999;6(2):104-14.
- [11] Lewis JR. *Ibm Computer Usability Satisfaction Questionnaires - Psychometric Evaluation and Instructions for Use*. *International Journal of Human-Computer Interaction* 1995 Jan;7(1):57-78.
- [12] Goud R, Peek N. *CARDSS' usability evaluation questionnaire*. Technical report TR2007-06, Dept. of Medical Informatics, Academic Medical Center, University of Amsterdam, 2007. Available at <http://kik.amc.uva.nl/KIK/reports/TR2007-06.pdf>. 2007.
Ref Type: Generic
- [13] Peute LW, Jaspers MW. The significance of a usability evaluation of an emerging laboratory order entry system. *Int J Med Inform* 2007 Feb;76(2-3):157-68.

Assessment of Biomedical Knowledge According to Confidence Criteria

Ines JILANI^a, Natalia GRABAR^a, Pierre MENETON^a and Marie-Christine JAULENT^a
^aINSERM, UMR S 872, Éq. 20, Les Cordeliers, Université Pierre et Marie Curie,
Université Paris Descartes, Paris, F-75006 France

Abstract. The characterisation of biomedical knowledge taking into account the degree of confidence expressed in texts, is an important issue in the biomedical domain. The authors of scientific texts use grammatical and lexical devices to qualify their assertions. We named these markers of qualification “confidence markers”. We present here the results of our efforts to collect confidence markers from full texts and abstracts, to classify them on the basis of semantics, and their use within our knowledge extraction system. We propose in this study, an implementation of these confidence markers for functional annotation of the human gene *Apolipoprotein (APOE)* thought to be involved in Alzheimer’s disease. As a result, we obtain, through the extraction system, triplets: (G, F, PMID), in which G is the gene *APOE*, F is its function found in texts and the PMID of the article from which this knowledge was extracted. Moreover, a spatial 3D of the triplets, relative to each other, is proposed depending on their respective confidence degree.

Keywords. Data acquisition- data capture, Data analysis-extraction tools, Modeling

Introduction

Bibliographical databases, such as Pubmed¹, are indexing increasingly large numbers of biomedical articles. They are essentially consulted by the medical, biological and bioinformatics communities, and the number of searches last year (2006-2007) rose to over 82.3 million, with 423 million page views².

This study follows on from the automatic extraction of knowledge about genes from scientific articles indexed in Pubmed, based on a natural language processing method [1]. Loss of context limits the extraction of information about genes, such as their functions, the diseases associated with them and their interactions. For instance, if knowledge arises from an experimental procedure or constitutes a reference to another article, then the context allowing the recipient of the extracted knowledge to associate it with certain reliability and confidence is missing. There is an important issue to formalise this kind of additional information, to weight the extracted knowledge and use it more confidently.

¹ <http://www.ncbi.nlm.nih.gov/sites/entrez>

² NLM Technical Bulletin - June 18, 2007 MLA 2007: NLM Online Users’ Meeting Remarks
http://www.nlm.nih.gov/pubs/techbull/mj07/mj07_mla_dg.html (last visited on Oct, 19th 2007)

Information on the validity of knowledge is often given by a specific linguistic device, sometimes called “hedge”, “modifier” or “qualifier”. These linguistic phenomena can be referred to as “confidence markers”. They belong to different grammatical categories - verbs, adverbs or adjectives - and qualify the author’s assertions in the article. Consider, for instance, the following sentences “Copper deficiency is a plausible cause of Alzheimer disease (AD). This hypothesis should be tested with a lengthy trial of copper supplementation” (from the abstract of the article with Pubmed Identifier 17928161). The terms underlined are markers, indicating qualifications used by the author for tacit weighting of the reliability of his claim. The word “hedge” was first used in this area in 1972 by Lakoff [2], who described hedges as “words whose job it is to make things more or less fuzzy”. Hyland [3], Light [4], Mercer [5] later carried out qualitative studies of these qualifiers. However, they have neither modelled them, nor integrated their use for weighting any kind of information in a knowledge extraction system. In this study, we worked on the use of these confidence markers in scientific articles, their significance, their classification and their automatic detection in texts for knowledge weighting purposes. The main aim was to document the information so that it could be used confidently.

We first present the materials and methods used and implemented for our study, then the results obtained followed by their discussion and finally, a conclusion to place our results in their context and to describe the avenues for a further exploration.

1. Materials and Methods

Corpora. We used three corpora obtained by querying Pubmed [6]. CORP1 was obtained with a list of 160 human candidate genes thought to be related to AD [7]. This corpus is a collection of 355 abstracts, containing 817 sentences, 213,618 words and it is 1 MByte in size. CORP2 is composed of 68 full texts related to the abstract references in CORP1, available from Pubmed Central³. It contains 27,912 sentences, 1,123,873 words and is 2.4 Mbytes in size. CORP3 is a collection of 348 abstracts collected from Pubmed with a query containing a list of 160 nematode genes identified by biologists as potentially linked to AD. CORP3 contains 825 sentences, 201,753 words and is 1 MByte in size.

Lexical resource. WordNet® [8] is a large lexical database of English: nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms, each expressing a distinct concept. This database was used in our study to enrich the extracted confidence markers, by identifying their synonyms.

Our method consists in collecting the largest possible number of confidence markers (from CORP1, CORP2, CORP3), enriching them using WordNet®, classifying confidence markers and defining other confidence criteria. This enabled us to establish a confidence marker model, which was applied to our system for extracting knowledge concerning genes.

Collection of confidence markers. We collected confidence markers manually, retaining every linguistic device suggesting that the authors were placing a qualification on their claims, the presence of the author in the text, hints in the text relating to the context of the information presented, and so on.

³ <http://www.pubmedcentral.nih.gov/>

Enrichment of the list of confidence markers. The confidence markers collected provide a partial view of possible linguistic devices used by authors in scanned documents. However, it is useful to extend these devices to morphosyntactic variations and their synonyms. We used WordNet® resources for synonymic extension.

Classification of confidence markers and definition of other confidence criteria. The markers must be classified into homogeneous subcategories for use within an automatic extraction tool and for knowledge weighting purpose. These subcategories are further graduated according to the confidence or discredit they add to the extracted knowledge. Other criteria should also be taken into account to weight the extracted information: the impact factor (IF) is one of these. It is a measure of the frequency with which the "average article" in a journal is cited over a given period of time. The IF for a journal is calculated over a three-year period, and can be considered to be the average number of times published papers are cited up to two years after publication.

Modeling confidence criteria. The last step of the method involved establishing a model of the final set of markers. This confidence marker model must be useable within an automatic tool for extracting functional knowledge from texts, together with related information about confidence. This method is based on regular expressions (or graphs) constructed manually, as described elsewhere [1].

Finally, we applied this model to the context of AD and, more precisely, to searches for knowledge relating to the apolipoprotein E gene (*APOE*), which is thought to be linked to AD in humans. We built a specific corpus (CORP_APOE) for the application of the model. This corpus was obtained by querying Pubmed with *APOE* gene and its synonyms. We considered all biological functions of *APOE*, extracted them with the knowledge extracting system to get triplets (*G*, *F*, *PMID*), where *G* is the gene *APOE*, *F* is its function and *PMID* is the Pubmed identifier of the article from which the knowledge was obtained. Afterwards, we detected all confidence markers potentially qualifying the knowledge extracted, weighted and placed the triplets in a 3D space according to the criteria described above.

2. Results and Discussion

In this section, we detail the results obtained for each step of the proposed method.

Collection of confidence markers. A list of 250 manually collected confidence markers was generated.

Enrichment of the list of confidence markers. Once collected, we extended the markers, using their morphological variants. We do that by searching lemmas of the confidence markers. For instance, we extended the identified confidence marker "previous study" to "previous studies" (adding plural or singular forms), and "remain unknown" to "remains unknown" (inflectional form of verb). This process increased the number of confidence markers listed to 478.

We also carried out a second type of extension in which synonyms of the extracted confidence markers were taken into account. For instance, we extended the confidence marker "we anticipate" to "we expect" (*expect* being a synonym of *anticipate* in WordNet®), "previous study" to "previous work" and "previous report" (*work* and *report* are synonyms of *study*). There is some debate concerning the use, sufficiency, accuracy and linguistic relevance of WordNet® [9] but it was considered appropriate for the context in which we were working. Indeed, the process of extension based on

synonyms used here did not modify outcomes, it simply increased the size of the set of confidence markers to 700. Notice that ignoring this step of acquiring variants and synonyms of confidence markers would have resulted in a loss of weighting information, as our list of markers would not have been sufficiently exhaustive.

Classification of confidence markers and definition of other confidence criteria. The necessity to classify the confidence markers according to their semantics has emerged in order to give different weights to the knowledge they characterise. We highlighted four different categories, described below in ascending order of confidence:

1. **Interrogation or trial and error of the author.** Knowledge that remains unproven and requires demonstration. This knowledge may also correspond to the author's interpretation in the absence of conclusive proof within the article.
e.g.: "remain to be confirmed", "has yet to be identified", "?"

2. **Distance suggested by the author compared to his assertions or the knowledge presented in the text.** This fine distinction was suggested by Hyland [3]. It may also correspond to a restriction of the knowledge concerned to a specific context (e.g.: the context of the article or experiment).
e.g.: "our findings suggest that", "in this case we conclude that", "it is possible that"

3. **Studies by other researchers, references to other works, articles or methods.** These elements are associated with a high level of confidence in the knowledge communicated, as we assume that if an article is cited, the information is assumed, or at worst simply believed to be true. This information therefore does not require further demonstration in the article concerned.
e.g.: "previous observation", "it is now believed that", "it has been proposed that"

4. **Demonstration or proof given by the author.** This corresponds to work carried out by the author and presented in the article concerned. It involves the use of markers implying that the knowledge is highly reliable and based on demonstrations presented within the article. These markers are associated with the highest level of confidence in knowledge.
e.g.: "we reveal that", "we show here that", "our results indicate that"...

These four categories correspond to *Type 1* confidence markers.

We also distinguished *qualifiers*, modifying confidence levels within the four categories described above. These qualifiers characterise subjectivity in texts, describing things or events that are possible but not observed or not certain. These qualifiers are *Type 2* confidence markers. They are represented in Fig. 1, from negation to affirmation, i.e. from the least probable to the most probable [10].

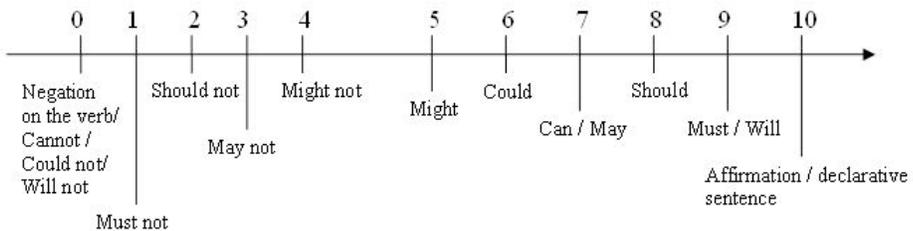


Figure 1. Graduation of *Type 2* confidence markers

Negation indicates that the author presents a negative assertion, with the author considering the knowledge negated to be false. This knowledge is therefore of the lowest confidence (“unable”, “not”, “neither...nor”...). *Must not, should not, may not* and *might not* are closer to negation than the other markers to the right of the arrow (*might, could,...*) and convey more confidence than a clear negation. *Might, could, can/may* and *should* express increasing probabilities. Finally, *must/will* and affirmative sentences express the highest positive degree of certainty.

Through the use of IF criterion, we hypothesised that the IF of a journal is related to the reliability of the biological and medical information published. This does not imply that journals with a low IF publish erroneous biological experiments, but simply that biologists are used to give more confidence weight to information derived from journals with a high IF. The publication of an article in a journal can be readily found using the ISSN (international standard serial number) of the journal, available in the XML format of the result of a Pubmed search, in the tag `<ISSN IssnType="Print">0003-2697</ISSN>` (the number in bold corresponds to the ISSN of the journal). A listing of the IF of journals can be used to weight the knowledge extracted from these journals.

Modelling confidence criteria for their automatic extraction. We used regular expressions (or graphs) [1] to retrieve information from texts and therefore modelled all the confidence markers in these terms. First, lemmas of each word in confidence phrases were defined, making it possible to detect all forms of a word, rather than just a single form. Second, synonyms of the markers, defined in the enrichment step, were collected into the same expression. For instance, confidence markers such as “we anticipate” and “we expect” form the expression $we\langle have \rangle * (\langle anticipate \rangle + \langle expect \rangle)$, where words between `< >` correspond to the canonical form of the word, `+` corresponds to logic *OR* and `*` means *0 or n* occurrences. When isolated markers were found to be synonyms within WordNet®, we included them in the same expression. This was the case for “we hypothesise” and “we suspect”, and, with extension for synonym, this created the following regular expression:

$we\langle have \rangle * (\langle hypothesise \rangle + \langle speculate \rangle + \langle expect \rangle + \langle predict \rangle + \langle suspect \rangle)$.

Nouns, adjectives, adverbs, as well as verbs, can be extended with WordNet® resource, as in the following expression:

$\langle have \rangle * \langle be \rangle (previously + now) * (largely + widely + extensively + generally) * \langle confirm \rangle$
which picks up the sentences “*have been previously confirmed*” as well as “*is now largely confirmed*” or “*is widely confirmed*”. Through these examples, we can observe the capacity of regular expressions to identify the targeted linguistic events.

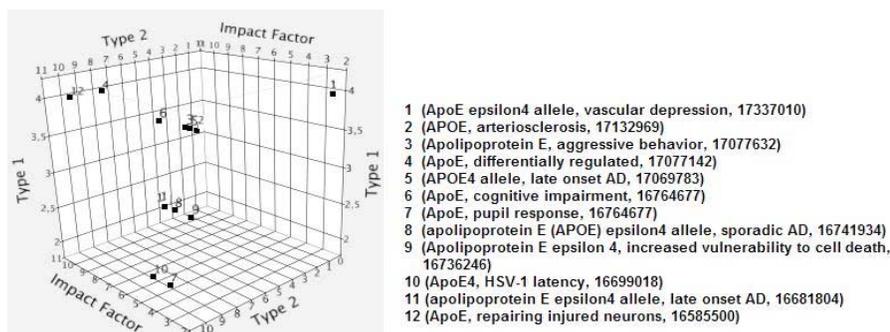


Figure 2. Graphical visualisation of the extracted triplets depending on their degree of confidence

Application in the context of apolipoprotein E gene. The application of such extended and precise regular expressions to CORP_APOE led to the extraction of 650 knowledge elements. These elements were extracted in the form of (*G, F, PMID*) triplets, where *G* is the gene, *F* its function and PMID the Pubmed identifier of the article from which the functional annotation was obtained.

This information can be visualised, as in fig. 2. The knowledge extracted is positioned in 3D space according to the *Type 1* confidence level of the biological knowledge, the *Type 2* confidence level and to the IF of the source journal.

3. Discussion and Conclusion

This paper describes our experience with the attempt to integrate confidence markers into an automatic knowledge extraction system [1] and the application of these markers to documents in the biological and medical domains. The markers were collected manually from corpora of abstracts and full scientific texts extracted from Pubmed. They were extended with the WordNet® resource and classified into four categories of *Type 1* or ten categories of *Type 2* depending on the semantics and strength with which they modify the knowledge reliability.

The extraction and characterisation of knowledge concerning the biological function of genes and their products, based on a system of this type, improves our understanding of the reliability of this information, particularly if the extracted knowledge is projected into a 3D space. We are currently working with biological experts to evaluate the efficiency of such a presentation.

We plan to work on the automatic detection of new confidence markers, to make the available list even more exhaustive. Such detection may be based on the syntactic category of elements within sentences and on their expected semantic role. Otherwise, we have observed that confidence markers may be clustered together within a sentence, and we will try to exploit this observation for their automatic detection. In addition, we plan to take into account the type of the study presented in the article, i.e. if it is an observational study (epidemiological), a controlled experiment, or a clinical essay.

References

- [1] I. Jilani, N. Grabar N, M.-C. Jaulent, *Fitting the finite-state automata platform for mining gene functions from biological scientific literature*. In: Proceedings of the Second International Symposium on Semantic Mining in Biomedicine, Jena, 2006
- [2] G. Lakoff, *Hedges: A study of meaning criteria and the logic of fuzzy concepts*, Chicago Linguistic Society, **8** (1972), 183-228
- [3] K. Hyland. *The author in the Text: Hedging Scientific Writing*. Hong Kong Papers in Linguistics and Language Teaching, 1995.
- [4] M. Light, X.Y. Qiu, P. Srinivasan, *The language of bioscience: facts, speculations, and statements in between*. In BioLINK 2004: Linking Biological Literature, Ontologies, and Databases. 2004, p. 17-24
- [5] R.E. Mercer, C. Di Marco, *A design methodology for a biomedical literature indexing tool using the rhetoric of science*. Linking Biological Literature, Ontologies and Databases, HLT-NAACL 2004 Workshop: Biolink 2004, Association for Computational Linguistics, 2004.
- [6] N. Grabar, M.-C. Jaulent, A. Chambaz, C. Lefebvre, C. Néri. Sifting abstracts from Medline and evaluating their relevance to molecular biology. *Stud Health Technol Inform* **124** (2006), 111-116.
- [7] C. Lefebvre, J.-C. Aude, E. Clément, C. Néri, *Balancing protein similarity and gene co-expression reveals new links between genetic conservation and developmental diversity in invertebrates*. *Bioinformatics*, **21**(2005):1550-1558.
- [8] C. Fellbaum, ed, *WordNet, An Electronic Lexical Database*, The MIT Press, Cambridge (MA), 1998
- [9] M. Slodzian, *WordNet: what about its linguistic relevancy?* in Proc. of the EKAW2002 workshop on "Ontologies and texts". Juan-les-Pins (F), 2000.
- [10] I. Jacobs, *English modal verbs*, New York, 1995

Using Knowledge for Indexing Health Web Resources in a Quality-Controlled Gateway

Michel JOUBERT^a, Stefan J DARMONI^b, Paul AVILLACH^{a,c}, Badisse DAHAMNA^b,
Marius FIESCHI^a

^a *LERTIM, Faculte de Medecine, Universite de la Mediterranee, Marseille, France*

^b *CISMeF, Centre Hospitalo-Universitaire de Rouen, France*

^c *LESIM, INSERM U593, ISPED, Universit  Victor Segalen, Bordeaux, France*

Abstract. Objectives: The aim of this study is to provide to indexers MeSH terms to be considered as major ones in a list of terms automatically extracted from a document. Material and methods: We propose a method combining symbolic knowledge - the UMLS Metathesaurus and Semantic Network - and statistical knowledge drawn from co-occurrences of terms in the CISMeF database (a French-language quality-controlled health gateway) using data mining measures. The method was tested on CISMeF corpus of 293 resources. Results: There was a proportion of 0.37 ± 0.26 major terms in the processed records. The method produced lists of terms with a proportion of terms initially pointed out as major of 0.54 ± 0.31 . Discussion: The method we propose reduces the number of terms, which seem not useful for content description of resources, such as “check tags”, but retains the most descriptive ones. Discarding these terms is accounted for by: 1) the removal by using semantic knowledge of associations of concepts bearing no real medical significance, 2) the removal by using statistical knowledge of non-statistically significant associations of terms. Conclusion: This method can assist effectively indexers in their daily work and will be soon applied in the CISMeF system.

Keywords. Knowledge-based systems; Terminology-vocabulary; Indexing.

Introduction

The dissemination of medical scientific knowledge to the general public, students and health professionals is today well and truly based on the Internet. In France and countries in Europe, North America, Africa and Asia, where French is spoken, there exist numerous web sites providing health-related information. Some sites deliver accredited information, the information delivered by the others is not. The former received their accreditation from recognized bodies such as the CISMeF gateway (acronym for Catalog and Index of French-language Health Resources on the Internet, the most important and quality-controlled source of institutional health information in French) [1], or the Health on the Net Foundation throughout the rest of the world [2].

In a system as CISMeF, for instance, the problem is to index a large corpus of health-related documents. Today, indexing scientific documents is conducted using generally the MeSH thesaurus. Researchers in this field have envisaged doing

automated indexing systems for French-language documents, which should achieve the same capabilities as those developed for English [3]. However, in the long term, other terminologies in addition to MeSH should be used in order to index documents other than of a pedagogical or scientific nature, such as clinical reports, laboratory examinations, etc. The VUMeF (Unified Medical French-language Vocabulary) project [4], sponsored by the French National Research Agency (ANR), has been working along these lines. Its ambitious objective was to integrate the various French-language terminologies into a single homogenous package. To this end, it has drawn on the results of another project, the UMLF (Unified Medical Lexicon for French) [5], also sponsored by the ANR, the aim of which was to produce an as-complete-as-possible French-language Specialist Medical Lexicon.

The aim of this study is to propose a method related to the development of tools designed to facilitate indexing a large documentary database in the health domain. The objective is to develop a complete automated indexing process involving two components: 1) an indexer that extracts MeSH terms from documents, and 2) a process that filters extracted terms and retains only the most informative ones. We describe in this paper the method we propose to implement this latter component. We use symbolic knowledge: the UMLS Metathesaurus and Semantic Network [6]. We also use statistical knowledge derived from co-occurrences of MeSH terms in the CISMef gateway documentary database. The method we propose is inspired by a model of documents coding/indexing we previously applied on coding standardized discharge summaries in a French hospital [7]. We illustrate here the application of the proposed method to a set of records provided by the CISMef gateway each containing a list of MeSH terms. This work is part of a wider project of semi-automated indexing of some documents in the CISMef database: a first process extracts automatically MeSH terms from documents, after what our method selects some of them according to their computed relevance, providing indexers with an effective help system. This study follows previous research on the automated indexing of French-language health-related web resources [8].

1. Material and methods

1.1. UMLS symbolic knowledge

The UMLS knowledge sources we used (2007AA release) are the Metathesaurus and the Semantic Network. The Metathesaurus provides an inventory of the concepts used in the health domain drawn from the nomenclatures included in the UMLS, more than a hundred today, including MeSH. Each MeSH term is thus connected individually to a concept in the Metathesaurus. Each concept is connected to at least one type of concept in the Semantic Network. Binary semantic relationships are defined between the different types of concepts conveying declarative medical knowledge such as “causes”, “diagnoses”, “prevents”, “complicates”, etc. These relationships thus make it possible to propose meaningful connections between MeSH terms.

In a previous study we shown how these relationships could be conveyed by qualifiers that are associated with the terms in order to build queries for a document server using MeSH for indexing purpose [9]. For instance, “a diagnostic procedure diagnoses a disease” associates the qualifier “diagnostic” with the term designating the disease. This knowledge thus enables us to propose associations of MeSH terms, some

of which are qualified, for comparison with the associations found in the list of terms in a record. Only pairs of terms having a semantic association issued from the Semantic Network are considered after this first step.

1.2. Statistical knowledge

Each resource is indexed in the CISMeF database with a set of MeSH terms. It is then possible to construct the entire set of term co-occurrences from the set of records and to associate a frequency to each pair of co-occurring terms. We thus obtain a set of distinct pairs of co-occurring terms and their frequency of occurrence.

The *confidence* (*Conf*) we have to find a term B when a term A is present is the ratio between the number of times that A and B, *Freq* (A,B), co-occurred and the number of times that A has occurred independently of the other terms, *Freq*(A):

$$Conf(A,B) = Freq(A,B) / Freq(A)$$

Confidence is the equivalent to the conditional probability of B if A in terms of probabilities [10]. It indicates the degree of interest of the association between A and B. However, confidence alone is not sufficient.

The *interest* (*Int*) we will have to consider the association of A and B is measured by the ratio between the confidence of the association of A and B and the frequency of B:

$$\begin{aligned} Int(A,B) &= Conf(A,B) / Freq(B) \\ &= Freq(A,B) / (Freq(A) * Freq(B)) \end{aligned}$$

Interest is the equivalent to *lift* as used in data mining and which is the coefficient by which we have to multiply the a priori probability of B in order to obtain the conditional probability of B, A being given. Interest measures the informational value of the association of A and B as compared with A and B taken separately. Thus, a high value of *Int*(A,B) means that A being given, the probability of B is high, and conversely.

Let us take the list of terms for a given record. Let us consider one of the terms that we will name A. The other terms in this same list will be named B_i. Knowing A, and the entire set of the ordered (A,B_i) pairs, we can obtain *Conf*(A,B_i) and *Int*(A,B_i) for each of the B_i. This allows us to assert, knowing A, which B_i are of interest. And conversely, knowing the B_i, we can suggest whether A is of interest or not. In both instances, it is necessary to determine the value of the interest above which the pairs of co-occurring terms are to be considered. This value is entirely heuristic and has an impact on the quality of expected results. It determines the terms to take into account in each record according of the relative interest they have to be associated with the other ones in the same record.

1.3. Material

Each CISMeF record contains MeSH terms, with or without qualifiers, and librarians according to indexing rules have designated certain terms as major terms. In order to test the above method, we built a set of co-occurrence pairs using CISMeF records and

applied it to the records returned by the CISMef gateway in response to 100 requests formulated by five different teams. The system replied successfully to 89 requests. We limited to 10 the number of responses to each request as this generally matches the first page of responses returned by a search engine (between 66% and 85% of expected results are present in the first page displayed [11]). This gave us a total of 293 successful answers.

Table 1 shows the number of different terms, the number of different major terms, and the number of co-occurrences between terms (major terms included) in the test sample. Co-occurrences were limited by adopting a lower threshold set at 0.5 and 1.0 for interest.

Table 1. Numbers of terms, major terms and co-occurrences for interest set at 0.5 and 1.0.

Number of terms	1,400
Number of major terms	740
Number of initial co-occurrences	9,242
Number of co-occurrences, interest 0.5	6,962
Number of co-occurrences, interest 1.0	4,460

2. Results

Table 2 shows the average numbers of terms, major terms and proportion of major terms per record in the test sample. The second column of the table presents these values for records indexed humanely. They will constitute our standard in what follows. Columns 3 and 4 shows the values obtained after applying the method with an interest value set at 0.5 and 1.0. The results show three things.

- the average numbers of terms decreases when the value of the interest increases
- the average number of major terms is not statistically different between filtered lists of terms and initial ones
- the average proportion of major terms in processed records increases with the value set to the interest.

Even if there are no statistically significant differences, we nevertheless try to interpret the results. A value of interest set to 1 is too restrictive because it decreases too much the average number of major terms. So, considering a value of interest set to 0.5, the average number of major terms is more or less the same than the standard value, and the proportion of major terms in the lists of processed records is higher than the standard value.

Table 2. Average number of terms, major terms and proportions of major terms per record in the test sample, and after applying the method with interest value set at 0.5 and 1.0

Interest		0.5	1.0
Average number of terms	8.17 ± 3.36	4.32 ± 2.41	3.55 ± 2.05
Average number of major terms	2.52 ± 1.41	1.94 ± 1.24	1.74 ± 1.11
Average proportion of major terms	0.37 ± 0.26	0.50 ± 0.31	0.54 ± 0.31

3. Discussion

We applied the combination of symbolic and statistical knowledge successfully in a previous work [7]. Even if the intent was not the same (coding discharge summaries on one hand, indexing documents on the other one), and the implementation of the model was different, they both exploit the symbolic knowledge of the UMLS and co-occurrences (of diagnoses and medical acts on one hand, of terms on the other one).

The joint use of both symbolic and statistical knowledge enables our proposed method to achieve an appreciable level of efficacy. Symbolic knowledge uses the UMLS Semantic Network to provide associations of medically significant concepts. The statistical knowledge built in a field of application allows taking into account a context for the symbolic knowledge and makes it possible to attribute a numerical value to the medically relevant associations. This signifies that it eliminates a number of terms deemed to be of lesser importance for the description of the documents content. Facing the challenge of a semi-automated indexing process of health documents, it retains the most descriptive ones. Values of Table 2 show that the average number of retained terms decreases when the value set to interest increases. This signifies that a small value of the interest is in favour of a number of terms greater than those retrieved with a high value of interest, due to the retained number of associations (Table 1). But, a small value of the interest makes the average number of major terms quite equal to the standard value. Lastly, we can remark that the number of records that the method has difficulties to fully process increases with the threshold set for the interest. So, we have to tune the value set to the interest according to the relevance of the method. In the case of this experiment our choice would be to set the value of the interest to 0.5. That allows discarding a notable number of terms while retaining the most of major ones.

Some terms, such as “check tags” (concepts of potential interest, regardless of the general subject content of the resource including terms such as *human*, *animal*, and *case-report*), having neither semantic nor statistical relations with other terms in records are discarded. Nevertheless indexers for resources indexing/retrieval purpose must introduce them. The method sometimes supplies a term with no qualifier and the same term with qualifiers. Let us consider, for instance, a record containing the following terms where the symbol «*» indicates a major term: *hematuria*; **hematuria/diagnosis*; *hematuria/etiology*; *urology/teaching and education*; *signs and symptoms*. After processing the list of suggested terms is: *hematuria*; *hematuria/diagnosis*; *hematuria/etiology*. The document in question deals with diagnosis of hematuria and therefore there are good grounds for indicating that “hematuria” and “etiology of hematuria” could be considered as significant terms. The method discards the terms *signs and symptoms* and *urology/teaching and education* that are important for retrieving purpose, but not significant for describing precisely the content of the document, as it does with “check tags”.

4. Conclusion

The proposed method aims to provide two-pronged assistance to human indexers in their work. Firstly, by using statistics to check that the terms indicated by an indexer comply with the established statistics. Secondly, in the light of the established statistics, by interactively submitting to indexers appropriate keywords and qualifiers according

to the keywords that he/she has already introduced into the record of the document currently being indexed. The method can also be used to check indexing quality by verifying terms introduced by an indexer and can then issue warnings when unlikely associations of terms occur.

The CISMeF team developed automated indexing tools since 2002 [3, 12]. Then it was decided to use them in the daily practice for most of the Internet resources. Three levels of indexing were defined: 1) totally manually indexed resources (e.g. guidelines); 2) resources automatically indexed and then reviewed by medical librarians; 3) totally automatically indexed resources (e.g. teaching material). The method proposed in this work will be soon applied to resources falling within levels 2 and 3. The use of the method is particularly interesting in the latter case. An automated indexation of documents may produce a lot of terms that the method can filter to retain only the most descriptive ones according to statistical and symbolic knowledge bases.

Acknowledgements

The authors thank the whole consortium of the VUMeF project in the framework of which this work was conducted. They also thank the U.S. National Library of Medicine for free access to the UMLS knowledge sources. They thank Mr George Morgan for his translation.

References

- [1] Darmoni SJ, Leroy JP, Baudic F, Douyère M, Piot J, Thirion B. CISMeF: a structured health resource guide. *Methods Inf Med.* 2000 Mar; 39(1): 30-5.
- [2] Boyer C, Selby M, Scherrer JR, Appel RD. The Health On the Net Code of Conduct for medical and health Websites. *Comput Biol Med.* 1998; 28(5): 603-10.
- [3] Neveol A, Mork JG, Aronson AR, Darmoni SJ. Evaluation of French and English MeSH indexing systems with a parallel corpus. *AMIA Annu Symp Proc.* 2005: 565-9.
- [4] Darmoni SJ, Jarousse E, Zweigenbaum P et al. VUMeF: extending the French involvement in the UMLS Metathesaurus. *AMIA Annu Symp Proc.* 2003: 824.
- [5] Zweigenbaum P, Baud R, Burgun A et al. Towards a unified medical lexicon for French. *Stud Health Technol Inform.* 2003; 95: 415-20.
- [6] McCray AT, Nelson SJ. The Representation of Meaning in the UMLS. *Meth Inform Med* 1995; 34: 193-201.
- [7] Avillach P, Joubert M, Fieschi M. A Model for Indexing Medical Documents Combining Statistical and Symbolic Knowledge. *Proc. AMIA Symp.* 2007. To appear.
- [8] Joubert M, Peretti AL, Darmoni SJ, Dahamna B, Fieschi M. Contribution to an Automated Indexing of French-language Health Web Sites. *AMIA Annu Symp Proc.* 2006: 409-13.
- [9] Aymard S, Falco L, Dufour JC, Joubert M, Fieschi M. Modeling and implementing a health information provider on the Internet. *Proc. MIE2003.* IOS Press; 2003: 89-94.
- [10] Geng L, Hamilton HJ. Interestingness measures for data mining: A survey. *ACM Computing Surveys* 2006; 38(3): art. #9.
- [11] Spink A, Jansen, BJ. *Web search: Public searching of the web.* Kluwer Academic Publishers 2004.
- [12] Névéal A, Pereira S, Kerdelhué G, Dahamna B, Joubert M, Darmoni SJ. Evaluation of a simple method for the automatic assignment of MeSH descriptors. to health resources in a French online catalogue. *Stud Health Technol Inform.* 2007; 129: 407-11.

On Machine Learning Classification of Otoneurological Data

Martti JUHOLA

Department of Computer Sciences, 33014 University of Tampere, Finland

Abstract. A dataset including cases of six otoneurological diseases was analysed using machine learning methods to investigate the classification problem of these diseases and to compare the effectiveness of different methods for this data. Linear discriminant analysis was the best method and next multilayer perceptron neural networks provided that the data was input into a network in the form of principal components. Nearest neighbour searching, *k*-means clustering and Kohonen neural networks achieved almost as good results as the former, but decision trees slightly worse. Thus, these methods fared well, but Naïve Bayes rule could not be used since some data matrices were singular. Otoneurological cases subject to the six diseases given can be reliably distinguished.

Keywords: Machine learning, Classification

Introduction

Vertigo or dizziness and other balance disorders are a common nuisance and can be symptoms of a serious disease. These problems are researched in otoneurology [1] to find their causes, to devise treatments and to pre-empt accidents originated from such harms. For this purpose, computational classification methods can be used to identify a patient's disease and separate disease cases from each other.

There are several classification methods using various approaches to group objects to different classes. Traditionally, discriminant analysis, cluster analysis, nearest neighbour searching, Naïve Bayesian rule and decision trees have been applied [2]. Afterwards, new methods have been developed such as neural networks [3]. It is infrequently, however, studied whether these methods have different potentials to classify various datasets. The aim of the current study was to define how different machine learning methods are able to classify otoneurological data. The general aim of this classification is to support otoneurological diagnostics. These diseases are often difficult for even experienced specialists to diagnose and distinguish from each other.

Earlier we investigated the classification of otoneurological data on the basis of multilayer perceptron neural networks [4,5]. We then encountered problems stemmed from a biased class distribution and a relatively small number of cases in the dataset. Small disease classes with a small number of cases were more or less lost in classification. In other words, large classes dominated the shares of the small classes in data distribution. These difficulties are fairly common in medical datasets: there are frequent and infrequent diseases. When the dataset included six diseases and their share

was not 1/6, but clearly greater or less, this biased class distribution made it difficult for perceptron neural networks to learn small classes. In principle, hidden nodes and thus a number of connections (weights) could be increased in a neural network to learn accurately enough the features of a dataset, but a relatively scarce number of the learning cases restricted this. To overcome these structural problems of the dataset, we later constructed a set of multilayer perceptron neural networks [5], a separate network for each disease class, when each network was trained employing both data of a disease and artificial data of its counterpart in variable space, but the process was complicated.

In the present study a comparison of machine learning methods was performed and, secondly, the data was preprocessed with principal component analysis (mainly for perceptron networks) to address problems caused by the structural features of the data.

1. Data

The dataset of 815 cases consisted of 38 attributes as replies to queries regarding patients' symptoms, medical history, results of laboratory measurements and clinical findings. These attributes were found to be the most important of a larger attribute set in our earlier research [6] and infrequently contained missing values. There were 11 % missing values in the whole dataset. They were imputed using modes for 11 binary attributes and one nominal attribute and using medians for 10 ordinal and 16 quantitative attributes. The only, four-valued nominal attribute was still substituted by three binary attributes to enable the use of Euclidean measure in its strictest and most reliable way, as this measure cannot be applied to nominal attributes, except binary ones. Imputation was carried out class by class being important to follow a class-wise distribution since it is, of course, essential that there are differences between classes, which is actually the footing of any classification. Using modes and medians in imputation is straightforward, but sufficient in this case, because, all in all, the number of the missing values was pretty small and we earlier observed [7] that the more sophisticated techniques of linear regression and Expectation Maximisation (EM) gave no better results while using discriminant analysis for the classification of otoneurological data. Another plausible cause was that the physicians who collected the data did not clarify some attributes seeing these less important for some certain cases.

The dataset was gathered at the Department of Otorhinolaryngology, Helsinki University Central Hospital, Finland. The six diseases and their frequencies are presented in Table 1. The subset of Menière's disease is far larger than two small subsets of sudden deafness and traumatic vertigo.

Table 1. Frequencies of the six diseases in the dataset.

Disease	Number	%
vestibular schwannoma	130	16
benign positional vertigo	146	18
Menière's disease	313	38
sudden deafness	41	5
traumatic vertigo	65	8
vestibular neuritis	120	15
total	815	100

2. Methods

The following methods were examined for the comparison: k -nearest neighbour searching, discriminant analysis, Naïve Bayesian decision rule, k -means clustering, decision trees, multilayer perceptron neural networks and Kohonen networks (self-organised maps). Nonetheless, data matrices required by the Bayesian method were inevitably singular since they incorporated such attributes that included purely zeroes for some classes and these were crucial variables impossible to eliminate. Singular matrices would rule out their inversions required. Thus, the Bayesian method was excluded. Correspondingly, discriminant analysis suffered from the similar complication in the situation of Mahalanobis (generalised Euclidean) or quadratic discriminant function. Therefore, only linear discriminant analysis was executed. All implementations were programmed with Matlab using various values of control parameters like numbers of clusters or network hidden nodes.

To test the dataset was divided into ten pairs of a learning subset (90 % of cases) and a test subset (10 %) in accordance with 10-fold cross-validation so that every case was incorporated in exactly one training set. The selection into the subsets was performed randomly, but in accordance with the class distribution. Sensitivity, specificity and total accuracies were computed. Most of machine learning methods involve random initialisations. Therefore, ten runs for each cross-validation pair and, thus, 100 runs altogether were repeated with an exception. Two of the test subset pairs were discarded for discriminant analysis, because they consisted of matrices, which were not positive definitive. Accordingly, 80 runs were executed for it. The results are shown in a condensed form in the following section.

3. Results

Tests of nearest neighbour searching were computed for one, three and five nearest neighbours. Their average results are presented in Table 2. An increase of nearest neighbours k did not improve otherwise satisfactory results which is possible depending on data. The values of vestibular schwannoma (acoustic neurinoma) and benign positional vertigo varied depending on k , which partially represents that their cases were sometimes mixed between the two classes. When the total accuracy value of each class was weighted with its number of the cases, mean weighted total accuracies were 93.5 ± 2.3 , 93.0 ± 2.9 and 93.2 ± 2.8 % for k equal to 1, 3 and 5.

Linear discriminant analysis executed is characterised in Table 3. This was exceptionally successful for the class of sudden deafness, which can be hard to distinguish medically on the one hand and being the smallest class on the other hand. The other classes were also detected efficiently. The mean weighted total accuracy of 95.5 ± 1.8 % was obtained.

Table 4 introduces results yielded by k -means clustering, where from several values of k three are given. The minimum was six because of the number of the classes. Apart from sudden deafness the number k of clusters between 12 and 40 produced approximately as high average results. For sudden deafness, 20 clusters sufficed to evolve sensitivities up to 89 %. Benign positional vertigo remained the poorest recognised class here. Mean weighted total accuracies of 90.9 ± 3.8 , 92.8 ± 3.5 and 92.9 ± 3.6 % were achieved for k equal to 6, 12 and 20.

Table 2. Means and standard deviations of nearest neighbour searching in percents for 100 test runs.

Number of neighbours	Static	vestibular schwannoma	benign positional vertigo	Menière's disease	sudden deafness	traumatic vertigo	vestibular neuritis
1-nearest	Sensitivity	88.5±9.8	80.7±19.7	87.6±6.5	92.5±12.1	82.6±19.7	86.7±13.1
	Total accuracy	96.4±1.6	91.8±3.6	90.4±4.8	99.4±1.0	97.8±2.2	96.6±2.1
3-nearest	Sensitivity	89.2±10.4	77.3±23.6	87.9±5.6	92.5±12.1	78.8±16.5	84.2±13.9
	Total accuracy	96.0±2.5	91.2±4.3	89.6±5.2	99.4±1.0	97.6±2.4	96.5±2.5
5-nearest	Sensitivity	79.2±16.2	83.3±18.9	89.5±5.8	90.5±12.3	80.2±16.4	90.0±11.7
	Total accuracy	96.4±2.7	92.2±3.4	89.1±6.0	99.4±0.9	97.9±1.7	97.1±2.2

Table 3. Means and standard deviations of linear discriminant analysis in percents for 80 test runs.

Static	vestibular schwannoma	benign positional vertigo	Menière's disease	sudden deafness	traumatic vertigo	vestibular neuritis
Sensitivity	87.5±8.6	87.5±15.5	89.6±4.8	100.0±0.0	90.5±13.3	95.8±8.4
Total accuracy	97.5±1.5	93.2±2.4	93.4±3.1	99.5±0.6	98.8±1.2	98.2±1.7

Table 4. Means and standard deviations of k -means clustering in percents for 100 test runs.

Number k of clusters	Static	vestibular schwannoma	benign positional vertigo	Menière's disease	sudden deafness	traumatic vertigo	vestibular neuritis
6	Sensitivity	81.2±14.2	66.2±33.4	92.8±7.3	12.0±32.7	53.8±44.4	84.8±20.2
	Total accuracy	93.7±3.2	88.1±6.0	88.0±6.7	95.5±1.6	95.3±3.4	94.9±5.1
12	Sensitivity	80.8±13.1	71.7±28.5	93.3±6.2	68.3±40.0	87.2±16.3	86.3±14.4
	Total accuracy	95.7±2.7	92.5±4.9	88.7±6.2	98.2±2.1	97.8±2.6	96.4±3.2
20	Sensitivity	81.2±13.3	74.9±26.0	91.2±8.1	88.8±17.5	85.7±18.4	87.2±13.6
	Total accuracy	96.0±2.4	92.3±5.4	88.6±6.4	99.3±1.0	97.9±2.2	96.8±2.2

Table 5. Means and standard deviations of decision trees in percents for 100 test runs.

Static	vestibular schwannoma	benign positional vertigo	Menière's disease	sudden deafness	traumatic vertigo	vestibular neuritis
Sensitivity	72.7±22.4	63.8±32.5	87.4±10.7	43.6±39.4	81.1±17.3	80.1±17.0
Total accuracy	95.3±3.5	89.0±4.5	82.6±4.8	94.9±2.4	96.6±2.1	95.4±3.1

Next decision trees (Table 5) were exploited pruning leaves to estimate the best tree size according to residual variance. The best size gained was 36.2 leaves on average. On average, decision trees were not effective for benign positional vertigo and

especially sudden deafness although their variances were large, i.e. some of trees were very good but others poor. A weighted mean total accuracy of $89.4 \pm 2.5\%$ was gained.

Multilayer perceptron networks as such were incapable to classify the dataset tending to put most cases to the largest Menière's disease class and to lose the two smallest classes entirely. Hence, principal component analysis was computed, results of which was used as input data. Every perceptron network comprised 40 input nodes (attributes) and six output nodes after the diseases. The only free parameter in the network topology was the number of hidden nodes. Adaptive learning and momentum coefficient were used with the backpropagation training algorithm, which was run no more than 200 epochs to prevent overlearning. Tests were conducted using 4-16 hidden nodes. The known recommendation states that the number of connections (weights) in a perceptron neural network should not be greater than one tenth of a training set so that efficient learning would succeed. Eight hidden nodes was such an upper bound, but the networks with 6-16 hidden nodes produced valid results (Table 6). Benign positional vertigo was the most difficult to detect. Mean weighted total accuracies of 92.9 ± 3.5 , 94.9 ± 2.9 and $95.0 \pm 3.0\%$ were achieved for 4, 6 and 10 hidden nodes.

Table 6. Means and standard deviations of multilayer perceptron networks in percents for 100 test runs.

Number of hidden nodes	Static	vestibular schwannoma	benign positional vertigo	Menière's disease	sudden deafness	traumatic vertigo	vestibular neuritis
4	Sensitivity	86.5±15.7	77.6±21.3	91.9±5.9	32.8±47.0	58.6±43.2	87.5±20.3
	Total accuracy	95.2±4.5	91.6±5.1	90.9±5.9	96.4±2.2	95.3±3.1	94.9±4.2
6	Sensitivity	89.3±9.6	79.0±20.8	91.53±5.5	93.2±23.4	86.9±16.7	91.9±10.9
	Total accuracy	97.7±1.8	93.0±4.1	91.7±5.4	99.3±1.2	98.2±1.6	97.3±2.4
10	Sensitivity	89.2±9.0	80.3±18.6	91.8±5.7	99.8±2.5	88.9±12.9	92.8±10.6
	Total accuracy	97.7±1.5	93.5±4.0	92.3±5.3	99.7±0.5	98.4±1.4	97.6±2.0

Table 7. Means and standard deviations of Kohonen networks in percents for 100 test runs.

Number of nodes	Static	vestibular schwannoma	benign positional vertigo	Menière's disease	sudden deafness	traumatic vertigo	vestibular neuritis
3×3	Sensitivity	78.5±17.2	66.1±27.2	94.1±6.6	32.2±28.5	13.9±27.6	84.5±16.5
	Total accuracy	94.3±2.8	86.6±4.9	87.2±7.0	93.9±2.3	91.6±2.9	96.2±2.8
5×5	Sensitivity	82.2±14.1	71.4±27.3	90.1±7.1	57.4±32.3	80.0±19.2	82.8±15.4
	Total accuracy	95.1±2.5	91.1±5.0	88.2±5.8	96.3±2.4	96.8±2.4	96.1±2.7
8×8	Sensitivity	86.2±12.9	74.6±24.5	88.1±7.7	79.4±26.6	84.8±18.2	84.6±15.4
	Total accuracy	95.9±2.5	91.5±5.2	88.7±5.9	98.4±1.8	97.7±2.3	96.4±2.6

Ultimately, Kohonen neural networks were assessed varying their sizes from 3×3 to 9×9 nodes. The latter was in the nature of the maximum, since there were 81 or 82 cases in each training set. Hexagonal neighbourhood pattern was employed as the

topological structure with link distance. For every network 400 learning epochs were completed. According to the results, 7×7 nodes were sufficient for the present data and as small as 5×5 for the other than sudden deafness. Table 7 includes a breakdown for three network sizes. Benign positional vertigo and sudden deafness obtained slightly poorer results compared to the remaining four classes. Mean weighted total accuracies of 90.3 ± 3.7 , 92.1 ± 3.5 and 92.7 ± 3.3 % were attained for the three sizes.

The running time (100 tests with a 2.6 GHz processor) of discriminant analysis was approximately 10 s. It was about 100 s for clustering and nearest neighbour searching, 5.5 min for decision trees and multilayer perceptron networks and 26 h 8 min for Kohonen 8×8 networks.

4. Discussion and conclusion

By adopting the best results from Tables 2-7, linear discriminant analysis was best for the current data by correctly classifying with the mean weighted accuracy of 95.5 %, though the differences were small between the methods. After feeding the principal components computed from the data, the perceptron neural networks of one hidden layer with 10 hidden nodes reached the next highest level of 95.0 %. Moreover, nearest neighbour searching of 1-neighbour achieved 93.5 %, k -means clustering of 20 clusters obtained 92.9 %, Kohonen neural networks of 8×8 nodes produced 92.7 % and decision trees 89.4 %. There were statistically significant differences between the first three mentioned and between them and the others according to t test (0.001 significance level). The running times favoured all excluding the Kohonen networks. On the other hand, the training phase was its time consuming part, which is only once accomplished or infrequently updated for practical purposes.

Other methods than multilayer perceptron networks did not benefit from using principal components as input. Perhaps this came from situation that the results were very good even without them. In fact, for linear discriminant analysis the results would be identical independent of this preprocessing for the sake of its linear character related to the principal component procedure. The results cannot be generalized for arbitrarily taken medical datasets. Still, they showed how the simple discriminant analysis was highly effective in accuracy and time.

References

- [1] J. Furman (ed.), *Otoneurology*, in: Neurologic Clinics, vol. 23-2, Elsevier, Netherlands, 2005.
- [2] A.R. Webb, *Statistical Pattern Recognition*, 2nd edition, John Wiley & Sons, Chichester, England, 2002.
- [3] K.J. Cios, W. Pedrycz, R.W. Swiniarski and L.A. Kurgan, *Data Mining, A Knowledge Discovery Approach*, Springer Science+Business Media, New York, NY, USA, 2007.
- [4] M. Juhola, K. Viikki, J. Laurikkala, I. Pyykkö and E. Kentala, On classification capability of neural networks: a case study with otoneurological data, in: Proc. of Medical Informatics 2001, London, UK, IOS Press, Amsterdam, Netherlands, 2001, 474-474
- [5] M. Siermala, M. Juhola, J. Laurikkala, K. Iltanen, E. Kentala and I. Pyykkö, Evaluation and classification of otoneurological data with new data analysis methods based on machine learning, *Information Sciences* **177** (2007) 1963-1976.
- [6] E. Kentala, I. Pyykkö, Y. Auramo and M. Juhola, Database for vertigo, *Otolaryngology – Head and Neck Surgery* **112** (1995) 383-390.
- [7] J. Laurikkala, E. Kentala, M. Juhola, I. Pyykkö and S. Lammi, Usefulness of imputation for the analysis of incomplete otoneurological data, *International Journal of Medical Informatics* **58-59** (2001) 235-242.

Semantic Relation Mining of Solid Compounds in Medical Corpora

Dimitrios KOKKINAKIS

*Department of Swedish Language, Språkdata
University of Gothenburg, Sweden*

Abstract. In the context of scientific and technical texts, meaning is usually embedded in noun compounds and the semantic interpretation of these compounds deals with the detection and semantic classification of the relation that holds between the compound's constituents. Semantic relation mining, the technology applied for marking up, interpreting, extracting and classifying relations that hold between pairs of words, is an important enterprise that contribute to deeper means of enhancing document understanding technologies, such as Information Extraction, Question Answering, Summarization, Paraphrasing, Ontology Building and Textual Entailment. This paper explores the application of assigning semantic descriptors taken from a multilingual medical thesaurus to a large sample of solid (closed form) compounds taken from large Swedish medical corpora, and determining the relation(s) that may hold between the compound constituents. Our work is inspired by previous research in the area of using lexical hierarchies for identifying relations between two-word noun compounds in the medical domain. In contrast to previous research, Swedish, as other Germanic languages, require further means of analysis, since compounds are written as one sequence with no white space between the words, e.g. *virus diseases* vs. *virusjukdomar*, which makes the problem more challenging, since solid compounds are harder to identify and segment.

Keywords. Terminology-vocabulary; Thesaurus-tools; Data-acquisition-data capture; Concept-representation; Knowledge-based systems

1. Introduction

Large text corpora in electronic form, the development and availability of large semantic knowledge bases, thesauri and other lexical resources as well as the lack of “intelligent” tools and techniques for semantically processing the data, have stimulated an ever increasing interest in the automatic acquisition of semantic relations even in the (bio-) medical domain. Semantic relations span a broad spectrum with semantic roles and semantic arguments, which usually are verb-centred, on the one edge, and pairwise semantic relations between compounds, usually noun compounds, on the other hand ([1][2][3][4][5]). The success of semantic approaches to relation mining are highly dependent on the structure, quality and coverage of the underlying lexical resource (ontology, thesaurus) and the linguistic pre-processing of the texts analysed. In this context, the identification and mapping of terminology from domain-specific literature, such as (bio-) medicine, onto a concept hierarchy constitutes a challenging and

important first stage of the semantic analysis of textual documents ([6][7][8][9]). In our work, we are particularly keen to identify the semantic relations of noun compounds in the (bio-) medical domain, and thus we only present relevant work related to this particular aspect of the problem. For this purpose, we explore the use of a large lexical hierarchy (MeSH) and a large Swedish medical corpus (MEDLEX).

In [10] noun-noun compounds (NCs) in the medical domain are classified using a set of 18 classes that describe the semantic relation between the head noun and the modifier in a given noun-compound (e.g. *purpose*, *material*, *cause*). The authors showed that a discriminative classifier can work quite well at assigning relations from a pre-defined set, if training data is supplied in a domain-specific setting, 60% accuracy is reported. In later work, [11] classified NCs using a multi-level hierarchy of semantic relations, with 15 classes at the top level. They provided a semi-supervised approach for characterizing the relation between two nouns in a biomedical NC based on the semantic category (MeSH) each of the constituent nouns belongs to. For instance, for the compounds *mesentery artery* and *leg vein* the first word of the NC falls into the category MeSH A01 (Body Regions), and the second word falls into A07 (Cardiovascular System), therefore such NCs pairs are hypothesized to express the same relation. By contrast, for the category pair A01-M01 (Body Regions–Persons) a distinction is needed between different kinds of persons and the algorithm needs to descend one level on the M01 side: M01.150 (Disabled Persons), M01.643 (Patients), M01.898 (Donors) etc. Thus, for *influenza patients* the first word falls under C02 (Virus Diseases) and the second under M01.643 (Patients). From these [11] declare that the relation that holds between the C02-M01.643 pair is *afflicted-by*. Their “descent of hierarchy” approach achieved a precision of 90% for finding the correct level of generalization, but it did not assign names to the relations. In [6] context features have been investigated in a way they can be applied for the identification of medical semantic relations. The authors designed a methodology to model relations and determine the parameters that distinguish these relations. The authors’ test was limited to only two relations, *treats* and *location-of*, but their claim is that the approach can be generalized to other domain-specific relations.

Our work replicates the methodology proposed by [11] who used the juxtaposition of category membership within the lexical hierarchy to determine the relation that holds between pairs of nouns (in English) with the addition of predicate paraphrasing inspired by [12] in order to propose a list of verbal predicates as designators of the relationship, without at this stage explicitly naming it. The remainder of this paper is structured in the following way. Section 2, Material, provides a short description of the medical corpus we use for our experiments and the characteristics of the MeSH thesaurus, the lexical resource we apply for semantic labelling of noun constituents. Section 3, Method, presents the experimental annotation of Swedish medical corpora. Section 4, Results & Discussion presents a qualitative and quantitative view of the results and discusses the findings. Finally, Section 5 summarizes the report, drawing some conclusions on the work done so far and presents thoughts for future research.

2. Material

2.1 The MEDLEX Corpus

The MEDLEX Corpus is the first large linguistically annotated Swedish medical corpus, [13]. The MEDLEX Corpus consists of a variety of text-documents related to

various medical text subfields, and does not focus at a particular medical genre, mainly due to the lack of very large Swedish resources within a particular specialized area. All text samples (20 mil. tokens, 48,000 documents) have been fetched from heterogeneous web pages during the past years, including: teaching material, guidelines, official documents, scientific articles from medical journals, conference abstracts, consumer health care documents, descriptions of diseases, definitions from on-line dictionaries, editorial articles, patient's FAQs and blogs etc. All texts have been converted to text files, and have been both structurally and linguistically annotated.

2.2 MeSH®

MeSH is the controlled vocabulary thesaurus (hierarchical terminology) of the NLM, U.S. National Library of Medicine. It is also a subset of the UMLS (Unified Medical Language System) Metathesaurus, the world's largest domain-specific thesaurus. The original data from NLM have been supplemented with Swedish translations made by staff at the Karolinska Institute Library based on the year 2006 MeSH. MeSH is used for subject analysis of biomedical literature, particularly for indexing the MEDLINE/PubMED, a large database of research papers from various sources. In MeSH each atomic or composite concept is assigned one or more alphanumeric semantic descriptor codes corresponding to particular positions in the hierarchy, so in that sense the thesaurus is *polyhierarchical*. However, the codes have a "transparent" structure and each place in the hierarchy is unique, from which both the semantic class of a concept and its depth in the hierarchy can be inferred, most general terms at the top, most specific at the bottom. For instance, the terms 'Kardiomyopati, hypertrofisk' *Cardiomyopathy, Hypertrophic* with the alphanumeric code C14.280.484.150.070.160 and 'Hjärtaneurysm' *Heart Aneurysm* with code C14.280.358, belong to the semantic group of "Cardiovascular Diseases" under node "C", but the node of the second term lies higher in the hierarchy as its code has fewer fields, three compared to six for the first term. There are 16 main subhierarchies in MeSH, e.g. 'A' Anatomy, 'B' Organisms, 'C' Diseases etc. See also: <<http://www.nlm.nih.gov/mesh/2007/MBrowser.html>>.

We utilize the terms from both the English and Swedish MeSH editions of 2006. We integrated the English hierarchy since it is fairly common that Swedish texts, intended both for professional and lay audience, contain portions of short or longer English segments, e.g. citations. Moreover, the use of English simplex or compound terms in Swedish texts is also common. This is probably due to the authors' unfamiliarity with the appropriate Swedish translation; by the influence or "contamination" from the English language, particularly in orthographic variation, e.g. use of 'ph' instead of 'f' (e.g. lymfo-lympho) and use of 'c' instead of 'k' (e.g. bradycardi-bradykardi); by spelling errors that might have a direct correspondence to English terms, the overuse of hyphen (e.g. tetra-cyklin instead of tetracyklin), or possibly because of an author finding the English spelling more appropriate.

3. Method

There are at least two partially interrelated aspects of the problem of the empirical interpretation of NCs that makes it particularly challenging and difficult. The first aspect, according to [14], deals with three components of the problem. That is, the actual identification of the compound, its syntactic analysis (left versus right

association) and the interpretation of the underlying semantics. Furthermore, the other aspect is tightly associated with the latter, and there are several reasons that contribute to its difficulty. [4] mention at least four factors: (i) NCs have implicit semantic relations (ii) NCs' interpretation is knowledge intensive and idiosyncratic (iii) there can be more than one semantic relation encapsulated in a pair of nouns and (iv) the interpretation of NCs can be highly context-dependent. Moreover, corpus-based semantic relation mining between terms stems from two main approaches: top-down or bottom-up. The first approach, which we also apply, consists in consulting an external semantic resource as a gold standard, and using this resource to determine if terms extracted from a corpus share a particular semantic relation, while the bottom-up approach is relying on statistical evidence ([15]).

We automatically annotated the MEDLEX corpus with the Swedish and English MeSH. Automatic compound analysis ([16]) determines the head and modifier part of the compounds which are then annotated accordingly. From the annotated sample we select all *binary* solid compounds, in which *both* constituents have been assigned a MeSH annotation, e.g.:

- <mesh id="C01...">tbc</mesh>-<mesh id="D24...">vaccin</mesh>; *TBC vaccine*
- <mesh id="C23...">blödning</mesh>s<mesh id="C23...">sjukdom</mesh>; *bleeding disease*
- <mesh id="B06...">kakao</mesh><mesh id="D03...">flavonoler</mesh>; *cacao flavonols*

If a head and/or modifier fell under more than one MeSH IDs, we made multiple versions of this categorization. In a N_1N_2 compound, N_2 is usually the head and N_1 the modifier. For instance, in the case of *andningsmuskler* i.e. *respiratory muscles*, we get the MeSH-annotation:

- <mesh id="G09...">andning</mesh>s<mesh id="A02.633;A10.690">muskler</mesh>

In which A02.633 stands for Musculoskeletal System and A10.690 stands for Tissue. In this case we make two versions of the annotated NC, namely:

- <mesh id="G09.772...">andning</mesh>s<mesh id="A02.633">muskler</mesh>
- <mesh id="G09.772...">andning</mesh>s<mesh id="A10.690">muskler</mesh>

There were a total of 85,000 compounds having two annotations¹, of which 15,000 were unique and 46,000 were the final number after making the multiple versions. Having identified and extracted the MeSH labelled NCs the next step was to recognize binary relation(s) that can exist between these semantically labelled NCs. It should be noted however, that there is currently no consensus as to which set of relations should hold between nouns in a noun compound, but most existing approaches make use of a set of a small number of abstract relations, typically between 10 to 50. To constrain the scope of the task, we have only chosen to examine a set of relations involving the MeSH category 'A', *Anatomy*, but the methodology should be extendable to all other combinations of the semantic categories. The number of compounds with an 'A' descriptor, in either the head and/or modifier position, was 11,423. The number of groups in which 'A' was at the modifier position was 10,452 (e.g. A+A 1505, A+B 158, A+C 3171, A+D 1404 etc.), while the number of groups in which 'A' was in the head position was 2,476 (e.g. B+A 91, C+A 262, D+A 282, E+A 83 etc.).

One way to characterize the semantic relationships that occur between the noun constituents is to discover the set of verbal predicates that paraphrase a NC. It has been shown [12] that this a promising approach, at least for English, compared to other forms of shallow semantic representations, such as preposition paraphrasing (i.e. N_2 preposition N_1 , as a paraphrase of N_1N_2) proposed by [14] and further improved by

¹ Note that "lexicalised" NCs are not taken into consideration in the experiment. This means that solid compounds in MeSH, such as 'bröstcancer' *breast cancer*, are excluded from the compound analysis and further semantic processing.

[17]. By applying paraphrases guided by a verbal predicate more finer-grained semantic representations can be compiled and captured. In order to achieve this goal the N_1 is paraphrased as a post-modifying relative clause, e.g. *bleedings disease* as: *disease that can-cause bleeding* or *disease that involves bleeding*. Following the methodology proposed by [12] we extracted sentences from the MEDLEX Corpus containing both N_1 and N_2 annotated them with part-of-speech information and filtered out relevant contexts based on the following pattern in which N_2 VERB N_1 are the only obligatory constituents:

N_2 REL-CLAUSE-INDICATOR AUX* VERB PARTICLE? PREP? ARTICLE? ADJECTIVE(S)* N_1
 Predicates occurring more frequent with for instance A+infection (e.g. hudinfektion *skin infection*) were: drabba *affect*, ta sig in *get into* and a large number of locative verb constructions such as ligga på *is/lies on*, sitta på *be on* and sitta i *is/lies in*, while predicates occurring more frequent with for instance A+bakteria *bacteria* (hudbakterie *skin bacteria*) were: finns på *occurs/exists*, infektera *infect*, förekomma *appear*, infiltrera *infiltrate*, tillhöra *belong*, befolka *inhabit*. Preliminary analysis showed that predicates that appear in combinations across the various semantic descriptors (e.g. A+C, B+E, D+M etc.) do not usually have any substantial overlap between their members. For instance the set of predicates for B+infection (e.g. bakterieinfektion *bacterial infection*) were: för/orsakas av *caused by*, bero på *depends on*; infektera med *infect with*; spridda via *spreading with*.

4. Results and Discussion

In the lack of suitable evaluation material, we manually determined if the identified verbs characterise the relationships between the descriptors at level-1 (A01, A02, A03 etc.) or there was a need to descent at lower levels, e.g. level-2 (A01.176, A01.236, A02.378 etc.) under the hypothesis that the members of two distinct level-groups fall into the same relation. This was necessary for very few categories in the examined set of NCs, such as ‘M’ *Persons* where the *descending* of one level was necessary (njurläkare *kidney physician* vs. njurpatient *kidney patient*), e.g.

<mesh id="A05...">njur</mesh><mesh id="M01.526...">läkare</mesh>

<mesh id="A05...">njur</mesh><mesh id="M01.643">patient</mesh>

There were many cases in which a paraphrase, as previously described, could not identify any relevant predicates between the paraphrased constituent even for NCs that this was clearly thinkable, e.g. spädbarnsinfektioner *neonatal infections*. This is drawback for the approach and in the future we plan to use directly the Web as a source of more material and even apply other paraphrasing models. There were also some problematic cases of NCs in which the head and/or modifier were referring to non-medical senses. This was particularly frequent with the noun huvud *head/main* as modifier e.g. huvuddiagnos *main diagnosis* annotated as ‘A01.456+E01’. The majority of the category pairs in hierarchy ‘A’ could fit into discrete semantic relations with overlapping predicate sets. For instance ‘A01+A02’ (ryggmuskel *back muscle* armlid *arm joint*) with predicates: ligga *lying*, sitta *sitt*, finns *be*, stabilisera *stabilize*), ‘A01+C17’ (ansiktsporiasis *face psoriasis*, handeksem *hand eczema* with predicates: förekomma *appear*, drabba *affects*, finns *be*), ‘D12/D14+A11’ (insulincell *insulin cell*, dopamincell *dopamine cell*) with predicates: bilda *build*, tillverka/produceras *produce*, insöndra *secrete*) and A02+C23 (muskelvärk *muscle pain*, skelettsmärta *skeletal pain*) with predicates: sitter *sitt*, uppkomma *arise/originate*, bero på *depends on*.

5. Conclusions

In this paper we have experimented with an idea of applying the juxtaposition of category membership within the MeSH thesaurus to determine the relation that holds between pairs of nouns in solid Swedish compounds, without explicitly labelling the relation itself. Exploring MeSH descriptors has the effect that the data becomes less sparse and thus more feasible to compute statistics over occurrences of descriptor codes than over words. In addition, predicate paraphrasing is used in order to propose a list of words as designators of the semantic relation(s) of the constituents. Our work replicates the methodology proposed by [11&12] this time on Swedish which poses restrictions w.r.t. the way the compounds are constructed. Compared to English, Swedish medical content on the Internet is quantitatively less, but in the near future we intend to explore that option, than merely a sampled corpus, for support of a larger set of predicates. Automatic extraction of semantic relationships between instances of a semantic resource is necessary in order to attach richer semantic metadata to documents than it is currently possible and can be useful for a broad range of language technology applications. In the future we intend to actually *name* the relations and experiment with a classifier construction for automating the process of assigning the semantic labels.

References

- [1] D. Gildea and D. Jurafsky. Automatic Labeling of Semantic Roles. *Computational Linguistics*. (2002). 28(3):245–288.
- [2] G. Leroy, H. Chen and J. D. Martinez. A Shallow Parser Based on Closed-class Words to Capture Relations in Biomedical Text. *Biomedical Informatics*. Vol. 36(3). (2001). Pp. 145–158.
- [3] Y. Kogan, N. Collier, S. Pakhomov and M. Krauthammer. *Towards Semantic Role Labeling & IE in the Medical Literature*. Proc. AMIA Annu Symp. 410–414. 2005.
- [4] R. Girju et al. On the Semantics of Noun Compounds. *Comp. Speech & Lang.* 19(4). (2005). 479–496.
- [5] K.B. Cohen and L. Hunter. A Critical Review of PASBio's Argument Structures for Biomedical Verbs. *BMC Bioinformatics* 7. (2006). Suppl 3:S5doi:10.1186/1471-2105-7-S3-S5
- [6] Š. Vintar, L. Todorovski, D. Sonntag and P. Buitelaar. *Evaluating Context Features for Medical Relation Mining*. ECML/PKDD Workshop on Data and Text Mining for Bioinformatic. Croatia. 2003.
- [7] M. Krauthammer and G. Nenadic. Term Identification in the Biomedical Literature. *Biomedical Informatics*. Special issue: Named entity recognition in biomedicine. Volume 37(6). (2004). 512 – 526.
- [8] K. Crammer, M. Dredze, K. Ganchev, P. P. Talukdar and S. Caroll. *Automatic Code Assignment to Medical Text*. Biological, Translational and Clinical Language Processing. 129-136. Prague. 2007.
- [9] A. Névéol, A. Rogozan and S. Darmoni. Automatic Indexing of Online Health Resources for a French Quality Controlled Gateway. *Information Processing and Management*. Volume 42:3. (2006). 695-709.
- [10] B. Rosario and M. Hearst. *Classifying the Semantic Relations in Noun Compounds via a Domain-specific Lexical Hierarchy*. Proc. of Empirical Methods in Natural Language Processing. 82–90. 2001.
- [11] B. Rosario, M. Hearst and C. Fillmore. *The Descent of Hierarchy, and Selection in Relational Semantics*. Proc. of the 40th Annual Meeting of the Association for CL. 247-254. Penn. US. 2002.
- [12] P. Nakov and M. Hearst *Using Verbs to Characterize Noun-Noun Relations*. Proc. of the 12th International Conference on AI: Methodology, Systems, Applications. Bulgaria. 2006.
- [13] D. Kokkinakis *Collection, Encoding and Linguistic Processing of a Swedish Medical Corpus - The MEDLEX Experience*. Proc. of the 5th Languages Resources and Evaluation (LREC). Italy. 2006.
- [14] M. Lauer *Designing Statistical Language Learners: Experiments on Noun Compounds*. PhD Thesis, Maquarie University. Australia. 1995.
- [15] M.T. Cabré, A. Condamines and F. Ibekwe-SanJuan, Application-driven terminology engineering. *Terminology*. (2005). 11:1, 1–19. John Benjamins Publishing Company.
- [16] D. Kokkinakis, *A Framework for the Acquisition of Lexical Knowledge; Description and Applications*. PhD Thesis. p. 73–75. Språkdata, University of Gothenburg. 2001.
- [17] M. Lapata and F. Keller Web-based Models for Natural Language Processing. *ACM Transactions on Speech and Language Processing*. 2:1. (2005). 1–31.

Use of the C4.5 machine learning algorithm to test a clinical guideline-based decision support system

Jean-Baptiste LAMY a,1, Anis ELLINI a, Vahid EBRAHIMINIA a

Jean-Daniel ZUCKER a, Hector FALCOFF b and Alain VENOT a

*a Laboratoire d'Informatique Médicale et de Bio-informatique (LIM&BIO), UFR
SMBH, Université Paris 13, Bobigny, France*

b Société de Formation Thérapeutique du Généraliste (SFTG), Paris, France

Abstract. Well-designed medical decision support system (DSS) have been shown to improve health care quality. However, before they can be used in real clinical situations, these systems must be extensively tested, to ensure that they conform to the clinical guidelines (CG) on which they are based. Existing methods cannot be used for the systematic testing of all possible test cases.

We describe here a new exhaustive dynamic verification method. In this method, the DSS is considered to be a black box, and the Quinlan C4.5 algorithm is used to build a decision tree from an exhaustive set of DSS input vectors and outputs. This method was successfully used for the testing of a medical DSS relating to chronic diseases: the ASTI critiquing module for type 2 diabetes.

Keywords. Knowledge-based systems, Expert systems, Evidence-based guidelines, Decision support, Assessment-evaluation, Type 2 diabetes, Machine learning

Introduction

Clinical Guidelines (CG) are useful for physicians [1]. However, guidelines printed on paper are difficult to use efficiently during medical consultation [2]. This has led to the development of decision support systems (DSS) based on the CG [3]. The ASTI project in France provides an example of such a system. This DSS aims to improve therapeutic care for patients with chronic diseases, by helping physicians to take into account the recommendations expressed in CG [4]. ASTI includes a critiquing module — a rule-based system composed of a knowledge base and an inference engine [5, 6]. This module is automatically activated when the physician writes a drug prescription; it compares the treatment proposed by the physician with that recommended by the CG, and issues an alert if they differ. ASTI has been applied to several chronic diseases, including type 2 diabetes.

Before their use in real clinical situations, DSS such as ASTI must be extensively

¹ Corresponding Author: Jean-Baptiste Lamy, LIMBIO, UFR SMBH, Université Paris 13, 74 rue Marcel Cachin, 93017 Bobigny cedex, France; E-mail: jibalamy@free.fr.

tested to ensure their medical validity. Errors may be encountered at various levels: in the knowledge base, in the inference engine or in the system specifications. Few papers on the testing of medical DSS have been published, although knowledge-based system verification has been widely studied outside the medical domain. Preece [7, 8] distinguishes between two types of verification methods. *Static methods* do not require the DSS to be run. They involve the inspection of the knowledge base by an expert, or checking for syntactic errors, logical errors (e.g. unsatisfiable conditions) or semantic errors (e.g. a male patient being pregnant) in the knowledge base [9]. These methods may identify errors, but cannot ensure the total absence of errors [7, 10].

Dynamic methods involve the running of the DSS with a test base. The test base may be written manually, or using automatic methods aiming to identify the “most relevant” test cases [10, 11]. The intervention of a human expert is required to determine whether the responses of the DSS are satisfactory. These methods therefore cannot be used for the systematic testing of all possible test cases, as there are generally too many such cases for manual review by an expert.

We aimed to check the conformity of the ASTI critiquing module to the CG used to design it — the French CG for type 2 diabetes [12]. We present a new dynamic verification method for “rebuilding” the knowledge contained in the CG from an exhaustive set of test cases, using machine learning techniques to construct a decision tree. We applied this method to the ASTI critiquing module for type 2 diabetes, and present the results of a comparison, by an expert, of the generated decision tree with the original CG. Finally, we discuss the potential value of such a method and possibilities of applying this method to other DSS.

1 Methods

We propose a general verification method with three steps: (1) generation of an exhaustive set of possible input vectors for the DSS, and running of the DSS to determine the output for each input vector, (2) extraction of knowledge from the set of (input vector, output result) pairs by applying learning or generalization algorithms, and (3) comparison, by an expert, of the knowledge extracted in step 2 with the original source of knowledge (here, the CG).

1.1 Generating Input Vectors and Outputs

It is possible to generate an exhaustive (or almost exhaustive) set of input vectors by considering a set of variables expressing the various elements of input for the DSS, and generating all possible combinations of the variables' values. Continuous variables (e.g. glycosylated haemoglobin) are limited to a few values, corresponding, for example, to the threshold values expressed in the CG. Finally, the output associated with each input vector is obtained by running the DSS.

1.2 Building the Decision Tree

A decision tree is built from the input vectors and the associated outputs, using C4.5 [13], a reference algorithm in machine learning. Pruning must be disabled, to ensure 0% error in the tree. Factorization rules are applied to reduce the size of the tree: (1) if

all the children of a given node include the same element of a recommendation (*e.g.* a recommended drug treatment), this element of information can be included in the node and removed from its children, (2) if a variable can take several values leading to the same recommendations, the largest set of such values can be grouped together as “<other>”.

1.3 Comparing the Decision Tree with Clinical Guidelines

Finally, the decision tree is compared with the CG by experts. These experts should be medical experts briefly trained in the evaluation method and the reading of the decision tree. The experts must check that the treatments recommended by the tree conform to the CG, and should check the CG to ensure that none of the recommendations included in the CG are missing from the tree.

2 Results

The general method presented above was applied to the ASTI critiquing module for type 2 diabetes.

2.1 Generating Input Vectors

The inputs of the ASTI critiquing module are related to (a) the patient's clinical condition, (b) the patient's treatment history, (c) the new treatment proposed by the physician, and (d) efficiency and the tolerance of the current treatment. For type 2 diabetes, these inputs correspond to the variables shown in Table 1. Doses were taken into account, using the following rule: if the treatment proposed is the same as the current treatment, the dose is understood to be modified: reduced if the problem identified was poor tolerance, and increased if the problem identified was low efficiency. A few rules were used to eliminate unrealistic combinations. For example, if the current treatment is diet, there cannot be drug intolerance.

Combinations of the values given in Table 1 generated an almost exhaustive set of input vectors. Only cases in which the patient had more than one drug intolerance, and cases involving quadritherapies (which are never recommended by the CG) were excluded. This approach yielded 147,680 input vectors.

2.2 Building the Decision Tree

The data were initially processed so that the list of valid treatments was used as the classifying variable, rather than the critiquing module's output (*i.e.* conform, not optimal or nonconforming). The decision tree was then built using the C4.5 algorithm. Before the factorization rules described in the methods section were applied, the decision tree included 87 nodes. After the application of these rules, the final tree included only 60 nodes (see Figure 1).

For a patient with a BMI of 25 kg/m² and an HbA1c level of 7%, with a current treatment of diet plus metformin shown to be inefficient, the decision tree recommends a first-line treatment of diet + metformin (with a higher dose of metformin as the

Variable	Definition	Retained values
Diabetes discovery	Was diabetes discovered at an early or late stage ?	Early, late
BMI	Body mass index	$\leq 27 \text{ kg/m}^2$, $> 27 \text{ kg/m}^2$
HbA1c	Glycosylated hemoglobin levels	$\leq 6.5\%$, $> 6.5\%$ (*)
Current type of treatment	The type of the treatment currently administered	No treatment, diet only, monotherapy, bitherapy, tritherapy, single daily insulin treatment, fractioned insulin treatment
Current treatment	Treatment currently administered	19 possible values, e.g.: no treatment, diet, diet+metformin
Problem	The medical problem preventing the continuation of current treatment	Low efficiency, poor tolerance
Efficiency	Level of efficiency of the current treatment (meaningful only if the problem is low efficiency)	Partial, null
Poorly tolerated drug	The drug in the current treatment responsible for intolerance (meaningful only if the problem is poor tolerance)	Metformin, sulfonamide, glinide, alpha-glucosidase inhibitors, glitazone, very slow-acting insulin, delayed-action insulin
Proposed treatment	The new treatment proposed by the physician, which may or may not conform to the CG	(As for current treatment)

Table 1: The list of variables considered for type 2 diabetes, and the values retained. (*) Only two values were considered, as ASTI makes use of treatment efficiency to determine whether the treatment should be strengthened.

recommended treatment is the same as the current treatment, see section 3.1), diet + glinide, diet + sulfonamide, diet + metformin + glinide or diet + metformin + sulfonamide, and a second-line treatment of diet + glinide + alphaglucoSIDaseinhibitors (AGI), diet + metformin + AGI, diet + sulfonamide + AGI, diet + metformin + glinide, diet + metformin + sulfonamide or diet + metformin + glitazone.

For a patient with a BMI of 29 kg/m^2 and an HbA1c level of 6.8%, with a current treatment of diet plus metformin, and who does not tolerate the metformin, the decision tree recommends a first-line treatment of diet + AGI, and a second-line treatment of diet + glinide or diet + sulfonamide. However, decreasing the dose of the metformin is not recommended.

2.3 Comparing the Decision Tree with Clinical Guidelines

We asked two physicians to compare the decision tree with the original CG. The experts found the decision tree to be easily readable, although one of them was at first confused by the use of only two levels of treatment intention on the tree. The experts searched the tree for the therapeutic recommendations expressed in the CG, and found that none of them was missing. They also ensured that the various branches of the tree lead to the same prescription that the CG. Consequently, the experts considered the tree to conform to the CG.

```

+--problem = low efficiency: diet+metformin OR
| | (diet+glinide+AGI OR
| | diet+metformin+AGI OR
| | diet+sulfonamide+AGI)
| +--BMI <= 27: diet+glinide OR
| | diet+metformin+glinide OR
| | diet+metformin+sulfonamide OR
| | diet+sulfonamide OR
| | (diet+metformin+glitazone)
| +--BMI > 27: diet+metformin+glitazone OR
| (diet+glinide OR
| diet+metformin+glinide OR
| diet+metformin+sulfonamide OR
| diet+sulfonamide)
+--problem = poor tolerance:
  +--BMI <= 27: diet+glinide OR diet+sulfonamide
  | +--poorly_tolerated_drug = metformin: diet+AGI
  | +--poorly_tolerated_drug = <other> : diet+metformin
  +--BMI > 27: (diet+glinide OR diet+sulfonamide)
    +--poorly_tolerated_drug = metformin: diet+AGI
    +--poorly_tolerated_drug = <other> : diet+metformin

```

Figure 1: Part of the decision tree generated, limited to patients currently treated by monotherapy and with HbA1c levels > 6.5%. Treatments in brackets are recommended as second-line treatments only. Treatments on non-terminal nodes apply to all the child nodes. AGI = alpha glucosidase inhibitors.

3 Discussion and conclusion

The verification method proposed made it possible to test the ASTI critiquing module for type 2 diabetes in an almost systematic manner. It also gives an overview of the DSS reasoning. The knowledge extracted from the DSS was represented using rules and decision lists as alternatives to decision trees. Rules were more verbose than the tree. The decision list was a little more concise, but it was easier to compare the tree with the textual CG, due to the tree-like sectioning of the CG. Other algorithms could be used, *e.g.* for controlling the tree's optimality. We tried to insert errors by deleting arbitrary rules in ASTI's knowledge base; these errors were clearly visible in the decision tree and were found by the experts.

The method proposed here has several advantages. First, it permits almost systematic testing. ASTI was preliminary tested with a hand-written test based on 796 input vectors. The method used here involved about 150,000 input vectors — three orders of magnitude greater than initially tested. Second, this method tests not only the DSS knowledge base, but also the inference engine. The decision tree obtained is not a simple conversion of the rule-based knowledge base into a decision tree. Indeed, the inference engine of the ASTI critiquing module includes several hard-coded generic rules: *e.g.* if the patient tolerated a drug poorly in the past, it should not be prescribed again, unless its dose is decreased. The decision tree took into account the knowledge base, but also these generic rules, including any possible bugs in the inference engine.

Finally, the method proposed considers the DSS as a black box, and thus makes no assumptions about its internal functioning. This method could therefore be applied to many other DSS, independently from the DSS reasoning method: rule-based system, hard-coded rules, decision tree, neural network, or any other algorithm. All DSS based on a human-readable source of knowledge, such as a CG, are eligible (by opposition to

e.g. a DSS using the k-nearest neighbours algorithm), which includes most of medical knowledge based DSS. To apply our testing method, one should first identify the DSS inputs, generate an almost exhaustive set of input vectors, and run the DSS for each of them. Then, one should extract knowledge from the input vectors and output data, using C4.5 or another learning algorithm. Finally, one should ask a human expert to compare the extracted knowledge to the knowledge that was used to build the DSS.

Three problems might be encountered when applying this method. First, it can be difficult to generate an almost exhaustive set of the DSS's input vectors, in particular when the number of variables is very high, e.g. for hypertension CG, or when there are many continuous variables. However medical recommendations usually define threshold values that allow to discretize continuous variable easily, as we have done for diabetes type 2. Second, it might be impossible to generate in a reasonable time the outputs for all input vectors (DSS too slow or too many vectors). Third, learning algorithms might be unadapted to the knowledge used by the DSS. For example, rule conditions such as "if the patient has two or more risk factors from a given list..." cannot be learnt by C4.5. To solve this problem, one can use more sophisticated learning algorithms, e.g. based on description logic, or derive input variables from the other variables, e.g. the number of risk factors could be added as an input variable.

In conclusion, the ASTI critiquing module for type 2 diabetes was almost exhaustively tested in this study, using an original black-box dynamic verification method. This method appears generic and therefore applicable to other medical DSS.

References

- [1] J.M. Grimshaw and I.T. Russel. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet*, **342**(8883):1317–22, 1993.
- [2] J.C. Dufour, J. Bouvenot, P. Ambrosi, D. Fieschi, and M. Fieschi. Textual Guidelines versus Computable Guidelines: A Comparative Study in the Framework of the PRESUID Project in Order to Appreciate the Impact of Guideline Format on Physician Compliance. In proceedings of the AMIA Annual Symposium, pages 219–223, Washington, DC, 2006.
- [3] P.A. De Clercq, J.A. Blom, H.H.M. Korsten, and A. Hasman. Approaches for creating computer-interpretable guidelines that facilitate decision support. *Artificial intelligence in medicine*, **31**:1–27, 2004.
- [4] B. Seroussi, J. Bouaud, H. Dreau, H. Falcoff, C. Riou, M. Joubert, C. Simon, G. Simon, and A. Venot. ASTI: a guideline-based drug-ordering system for primary care. In *Medinfo*, volume 10, pages 528–32, the Netherlands, 2001.
- [5] V. Ebrahiminia, C. Riou, B. Seroussi, J. Bouaud, S. Dubois, H. Falcoff, and A. Venot. Design of a decision support system for chronic diseases coupling generic therapeutic algorithms with guideline-based specific rules. *Studies in Health Technology and Informatics*, **124**:483–488, 2006.
- [6] V. Ebrahiminia, C. Duclos, M.E. Toussi, C. Riou, R. Cohen, and A. Venot. Representing the patient's therapeutic history in medical records and in guideline recommendations for chronic diseases using a unique model. *Studies in Health Technology and Informatics*, **116**:101–6, 2005.
- [7] A. Preece. Micro-level knowledge management, chapter Evaluating verification and validation methods in knowledge engineering, pages 123–145. Morgan-Kaufman, 2001.
- [8] A. Preece. Validation of knowledge-based systems: The state-of-the-art in north america. *Journal of communication and cognition - Artificial intelligence*, **11**:381–413, 1994.
- [9] G. Duftschmid and S. Miksch. Knowledge-based verification of clinical guidelines by detection of anomalies. *Artificial intelligence in medicine*, **22**:23–41, 2001.
- [10] M. Shwe, S.W. Tu, and L.M. Fagan. Validating the knowledge base of a therapy-planning system. *Methods of Information in Medicine*, **28**(1):36–50, 1988.
- [11] A. Ellini, V. Ebrahiminia, A. Venot, and J. Nobécourt. Une expérience originale de test automatisé d'un système d'aide à la décision clinique basé sur les guides de bonnes pratiques. In *IC'2007*, 2007.
- [12] HAS. Traitement médicamenteux du diabète de type 2. Available at http://www.has-sante.fr/portail/display.jsp?id=c_459266, November 2006.
- [13] J. R. Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann, 1993.

Process Mining for Clinical Workflows: Challenges and Current Limitations

Martin LANG^{ab1}, Thomas BÜRKLE^a, Susanne LAUMANN^b, Hans-Ulrich PROKOSCH^a

^a*Chair of Medical Informatics, University Erlangen-Nuremberg, Germany*

^b*Dpt. of Image and Knowledge Management, Siemens AG Medical Solutions, Germany*

Abstract. Process mining is an emerging technology in the context of Business Process Management with the goal to derive process models from observed system behavior. The global goals are: to detect previously unknown process structures, to implement consistent process controlling which may involve computation of realistic cycle times and frequency of occurrence of process pathways, or to quantify the conformance to guidelines. We did a detailed hands-on evaluation and analysis of established process-mining approaches and assessed their abilities to cope with the challenges of clinical environments. None of the examined 7 approaches fulfilled all requirements, but 2 could be circle out, which are to some degree suitable for clinical process mining.

Keywords. Process Mining, Clinical Workflows, Evaluation

Introduction

Due to increasing financial pressure and DRG based reimbursement, executives and controllers in a clinical environment are interested in optimized clinical and administrative processes. Often however, there is no knowledge about those processes and typical process analysis in a hospital environment is time consuming and tedious. Process variations are common and without rigid clinical pathways and a thorough deviation analysis their occurrence and frequency remains unknown. Shifts in workflows may go unnoticed for weeks or months. Process Mining has the potential to provide this kind of information by deriving process models from observed system behavior, like e.g. system log files. Currently, a broad variety of approaches are available [1-13], but the characteristics of clinical processes are challenging current solutions. In this paper we present a detailed analysis of existing mining approaches in respect to their ability to derive process models from real-life clinical data.

¹ Corresponding Author: Martin Lang, Chair of Medical Informatics, Friedrich-Alexander-University Erlangen-Nuremberg, Krankenhausstraße 12, D-91054 Erlangen, Germany; E-mail: martin.lang@m-lang.de.

1. Method

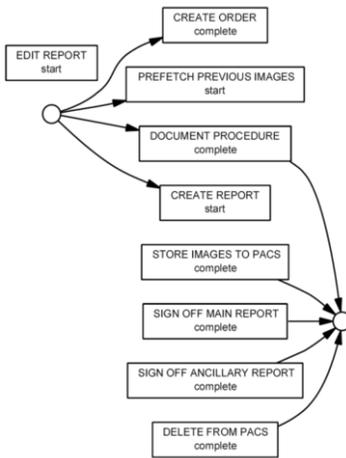


Figure 1. Mining Result of the α -algorithm using the CT exams log

To assess a selection of current process mining solutions we used data extracted from the log files of the radiology information system (RIS) and data of the hospital information system of the university clinic Erlangen, Germany. We examined event data from all ~15.000 Computer Tomography (CT), Magnetic Resonance Imaging (MR), ultrasound (US), and X-ray (XR) examinations performed in 2006. In this study we used the ProM-Mining Framework v4 [14] with 7 up-to-date algorithms. We analyzed and assessed the α -algorithm according to van der Aalst et al. [1], the $\alpha++$ -algorithm according to Wen et al. [2], the heuristic-mining algorithm according to Weijters et al. [3], the DWS-algorithm according to Greco et al. [4-5], the multiphase-algorithm according to van Dongen and van der Aalst et al. [6], the genetic-mining algorithm according to de Medeiros et al. [7], and

the theory-of-regions-based algorithm according to van Dongen and Busi et al. [8], as provided in the ProM tool. The study focused on the ability of the different approaches to derive correct process models from real-life clinical data and their ability to cope with the challenges of clinical processes. We chose clinical imaging as a research environment due to its high penetration with IT applications - able to deliver detailed logging information - and because clinical imaging processes compare well with other clinical processes. We analyzed and assessed the following criteria:

- Can the method deal with incomplete and noisy input data?
- Is it able to distinguish sequences, forks, and concurrency?
- Is it able to derive block-structured and arbitrary loops?
- Is it able to distinguish repeated activities from activities in loops?
- Can it deal with fuzzy process entry and end points?
- Is it able to detect different process types and variants?

To answer those questions, data was extracted from the log files and the database of the RIS, and transformed into a representation suitable for the mining algorithms. To assess the ability of an algorithm to induce complete and correct models, the *improved continuous semantics* fitness measure (ICS) [7] and the *source data verification* (SDV) methodology were used. The ICS fitness quantifies the ability of an algorithm to induce a model M covering the behavior given in a log L (completeness). It describes how good a process model covers the given behaviour. The ICS fitness assumes values between $(-\infty; 1]$ (1 is optimum) and is defined as follows with $f_p = a + b$ [7]:

$$ICS(L, M) = (|ParsedActivities(L, M)| - f_p) / |Activities(L)|$$

$$a = |MissingTokens(L, M)| / (|Instances(L)| - |InstancesWithMissingTokens(L, M)| + 1)$$

$$b = |TokensLeftBehind(L, M)| / (|Instances(L)| - |InstancesWithTokensLeftBehind(L, M)| + 1)$$

To assess the preciseness of mined models the SDV methodology was used. Therefore subgroups as e.g. routine CT or emergency CT were selected in an iterative refinement process and the corresponding mined models were analyzed. We checked if the mining

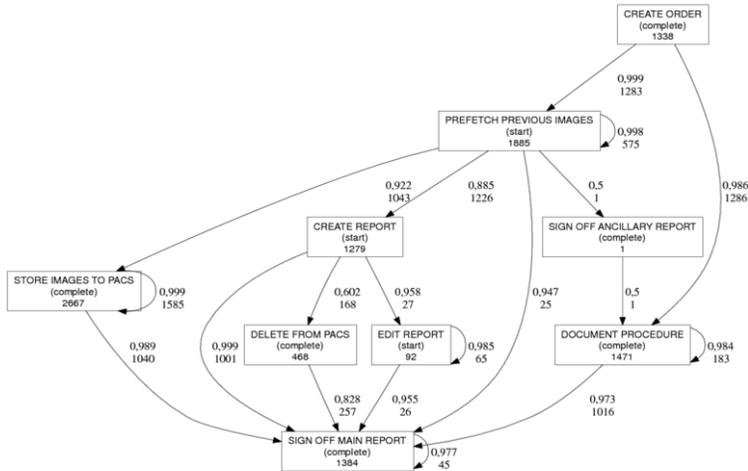


Figure 2. Mining Result of the DWS-Algorithm using the CT exams log data

would deliver artificial pathways. We e.g. choose a process with repeated activities to assess the capability of the algorithms to correctly distinguish repeated activities from activities in loops. The same course of actions was carried out to evaluate the other criteria like being able to deal with fuzzy process entry and end points, etc. If the induced process models allowed for artificial behavior, the corresponding algorithms were devaluated. A comparison of mined models with a reference process was not realized because binding reference processes are mostly missing, deviations are common and the behavior given in the logs is not necessarily complete.

2. Results

The study resulted in over 120 process models derived from the source data with and without filtering, selecting subgroups and so on. The details will be published in [15]. Here we present 2 typical examples for mining regular CT examinations with the α - (Figure 1) and the DWS-Algorithm (Figure 2). From those 2 models, which have been derived from identical source data, it becomes immediately clear, that some algorithms perform better, whereas others are not even able to detect a linear workflow of more than two activities. The DWS algorithm in Fig 2 produced a fairly good model, showing e.g. that from “Create order”, most cases will go to “Document procedure” and to “Prefetch previous images”. In the second case the workflow will continue in parallel with “Store images to PACS” and with “Create report” to finish with “Sign off main report”. In some cases however, images will be deleted from PACS or the report will be edited. For such a model we would now check if all direct connections between any two activities did exist at least for one patient in the source data. We obtained the following results: Models constructed using the approaches of Wen et al., Weijters et al., and Greco et al. (see also fig 2) showed only elements, which are true to the reality. The genetic miner of de Medeiros et al. – when used with heuristic initialization - did also produce correct models. When used with random-based initialization however, it showed artificial behavior under certain conditions. Though artificial elements will be penalized during the stepwise genetic evolution of the models, they cannot be ruled out

Table 1. Summary of the evaluation results (“+” for met, “-“ for unmet, and “+/-“ for partly met)

Assessment criteria	van der Aalst et al. (α algorithm)	Wen et al. ($\alpha++$ algorithm)	Weijters et al. (heuristic-miner)	Greco et al. (DWS-algorithm)	de Medeiros et al. (genetic algorithm)	van Dongen et al. (Multiphase miner)	Aalst and Rubin et al. (Region miner)
Truth to reality in contents	-	+	+	+	+/-	-	-
Noise and incompleteness	-	+/-	+	+	+	-	-
Sequences, forks, and concurrency	-	+	+	+	+	+	-
Loops	-	+	+	+	+	-	-
Repetitive activities	-	-	-	-	+/-	-	-
Fuzzy entry and end points	-	+/-	+	+	+	-	-
Process types and variants	-	-	-	+	-	-	-

completely, especially if only a few iterations are performed. We noticed artificial behavior also in the multiphase approach of van Dongen and van der Aalst, which e.g. would generate artificial process start and end points. The α - (see Figure 1) and the -region-miner-algorithm did generate none or incomplete models. The evaluation resulted in the following weighted and averaged ICS values for the mined CT, US, MR and XR models. -1,249 for van der Aalst et al., -1,305 for Wen et al., 0,683 for Weijters et al. and Greco et al., -0,733 for van Dongen et al., 0,919 for de Medeiros et al., and -1,294 for van der Aalst and Rubin et al. All algorithms except the α - and the region-miner-algorithm were able to derive forks, joins, as well as sequential and parallel behavior from the input data. Here the problem is that the α - and region-miner-algorithm require absolutely noise-free and complete input data, which cannot be assured when clinical data from log files is used, despite elaborate ETL (extraction, transformation and loading) methods used in this study. Four of seven algorithms ($\alpha++$ -, heuristic-, DWS- and genetic-algorithm) were able to correctly detect arbitrary and block-structured loops. One algorithm (van Dongen and van der Aalst et al.) derived loops, which did not correspond with the input data, whereas the α - and region-miner again did not produce sensible mining results in this task. The differentiation between repeated activities and activities in loops is an issue for the majority of the algorithms. Only de Medeiros et al. propose a variant of its genetic-mining algorithm, which is (partly) able to handle these constructs, by using the sets of successors and predecessors to distinguish between repeated activities and activities in loops. This implies a limitation to activities and excludes repetitive sub-processes, because the predecessors and successors within these sub-processes are not distinguishable from each other. All other investigated approaches were not able to handle these constructs, neither repeated activities, nor repetitive sub-processes. The efficient handling of fuzzy process entry and endpoints is another challenge for the approaches. Again, the α - and region-miner algorithm did generate insufficient results. In contrast, both heuristic- and DWS-miner perform well in this regard by counting successor frequencies globally. But this straightforward computation of relative successor frequencies (using activity

labels as identifiers) sacrifices the ability to distinguish repeated activities from activities in loops. This also holds true for the multiphase-algorithm. Besides, the latter generated artificial start tasks. The genetic-algorithm uses a guided evolution of potential process models to find the best fitting one for a given dataset. It tends to generate models representing the most frequent behavior of the input data and handles fuzzy process entry and end points well. Detection of different process types and variants was only possible with the DWS-algorithm. It uses a preprocessing step to cluster event traces based on string distance functions. The idea is, that process instances of the same type or variant have similar event traces. Actually, this preprocessing step is independent from the algorithm itself, but firmly integrated only in the DWS-algorithm. The other 6 algorithms failed on this task. A summary of the study results is given in Table 1.

3. Discussion

Analysis and evaluation of 7 different process-mining approaches revealed, that most mining algorithms have problems when analyzing event data from clinical workflows. This results either in failure to construct any valid process model or in models, which do not reflect reality. We could obtain correct (but partially incomplete and fractional) results only from 4 algorithms, namely Wen et al. ($\alpha++$), Weijters et al. (heuristic), Greco et al. (DWS), and de Medeiros et al. (genetic). One reason is that even clinical log data, which undergoes a thorough ETL process, is still to some degree noisy and incomplete, meaning that some data parts are missing or timestamps are incorrect because clock time might differ a fraction between modality and RIS. Thus algorithms, which operate with Boolean decisions to induce dependencies between activities, develop problems. One single negative data item in the log (e.g. caused by noise or incompleteness) will prevent correct detection of a dependency despite many hundred correct data items. In our study the α -algorithm, the multiphase-algorithm, and the theory-of-regions-based algorithm were badly affected by incomplete and noisy data. Surprisingly, the $\alpha++$ algorithm, which is also known to require noise-free and complete information, performed fairly well on our data set. We suspect this is due to the Boolean *isIndirectlyFollowedBy-Relation*, which is less susceptible to mine wrong or none edges in the presence of noise and incompleteness compared to the *isDirectlyFollowedBy-Relation* used e.g. in the α -algorithm. The heuristic-miner algorithm, the DWS-algorithm, and the genetic-miner algorithm produced good results despite noisy data. The challenge to correctly induce process models from clinical event data aggravates, if activities occur multiple times in the process models, without being part of a loop. In combination with noisy data, only the approach of de Medeiros et al. was capable of coping with this fact, but at the expenses of considerably higher computation times in the range of multiple hours for the data sets of this research. Only the approach of Greco et al. was able to detect process types and variants. This study does not cover all published process-mining approaches. The approaches from Cook and Wolf [9], Agrawal et al. [10], Pinter et al. [11], and Wen, Wang and Aalst et al. [12] were not considered, because of missing implementations or excessive information demands (e.g. requires explicit information about concurrency). Nevertheless, we think that the 7 algorithms tested in this study are a representative selection of available approaches, comprising the latest and most common approaches of the process mining research community. We conclude that none of the discussed approaches is able to

meet all major challenges of mining clinical processes such as noise, incompleteness, multiple occurrences of activities, or richness of process types and variants. The approaches of Weijters et al. or Greco et al. produced good results in the presence of noise and incompleteness combined with acceptable computation times in the range of minutes. Extending these approaches to enable them to cope with multiple occurrences of activities will be subject of future work. Apart from the fact that we considered only data taken from one hospital, almost every clinical information system provides such kind of logging data and therefore process mining can be applied. Even if ETL efforts, data granularity and quality may differ, the underlying characteristics of clinical workflows remain the same, regardless of the concrete infrastructure. Despite the mentioned flaws of current approaches, the concept of process mining carries great potential in helping to understand everyday clinical workflows and their variations. This will be helpful not only for process monitoring and controlling, but may eventually give us a better retrospective understanding of relations between diseases and corresponding treatment processes, once we are able to derive (partial) treatment process models merely by pushing a button.

References

- [1] W. van der Aalst, A. Weijters, L. Maruster: Workflow Mining: Discovering Process Models from Event Logs. *IEEE Transactions on Knowledge and Data Engineering*, 2004, 16(6), 1128-1142
- [2] L. Wen, J. Wang, J. Sun: Detecting Implicit Dependencies between Tasks from Event Logs. *APWeb 2006*, 591-603
- [3] A. Weijters, W. van der Aalst, A. de Medeiros: Process Mining with the Heuristics Miner Algorithm, Beta Working Paper, WP166, 2006
- [4] G. Greco, A. Guzzo, L. Pontieri, D. Sacca: Mining Expressive Process Models by Clustering Workflow Traces, *PAKDD*, 2004, 52-62
- [5] G. Greco, A. Guzzo, L. Pontieri, D. Sacca: Discovering Expressive Process Models by Clustering Log Traces, *IEEE Transactions on knowledge and data engineering*, 2006, 18(8), 1010-1027
- [6] B. van Dongen, W. van der Aalst: Multi-Phase Process mining: Aggregating Instance Graphs into EPCs and Petri Nets, 2nd PNCWB Workshop at the ICATPN, 2005
- [7] A. de Medeiros: Genetic Process Mining, PhD Thesis, University of Technology, Eindhoven, 2006
- [8] W. van der Aalst, V. Rubin, B. van Dongen, E. Kindler, C. Günther: Process Mining: A Two-Step Approach using Transition Systems and Regions, *BPMCenter.org*, 2006
- [9] J. Cook, A. Wolf: Automating process discovery through event-data analysis, *Proc of the 17th Intl Conference on Software Engineering (ICSE '95)*, ACM Press, 1995, 73-82
- [10] R. Agrawal, D. Gunopulos, F. Leymann: Mining Process Models from Workflow Logs, *Proc. of the 6th Intl Conference on Extending Database Technology (EDBT '98)*, Springer-Verlag, 1998, 469-483
- [11] M. Golani, S. Pinter: Generating a Process Model from a Process Audit Log. *BPM'03*, 2003, 136-151
- [12] L. Wen, J. Wang, W. van der Aalst, Z. Wang, J. Sun: A Novel Approach for Process Mining Based on Event Types, *IEEE Computer Society*, 2007, 00, 721-722
- [13] A. de Medeiro, W. van der Aalst, A. Weijters: Workflow Mining: Current Status and Future Directions, *On The Move to Meaningful Internet Systems*, Springer, 2003, 2888, 389-406
- [14] B. van Dongen, A. de Medeiros, H. Verbeek, A. Weijters, W. van der Aalst: The ProM framework: A new era in process mining tool support, *ICATPN*, 2005, 444-454
- [15] M. Lang: Prozess-Mining und Prozessbewertung zur Verbesserung klinischer Workflows im Umfeld bilderzeugender Fächer. PhD Thesis, University Erlangen-Nuremberg, Germany (in submission)

Searching Related Resources in a Quality Controlled Health Gateway: a Feasibility Study

Tayeb MERABTI^{a, b}, Suzanne PEREIRA^{a, b, c}, Catherine LETORD^a, Thierry LECROQ^a, Michel JOUBERT^b, Badisse DAHAMNA^a, Stéfan J. DARMONI^a

^a *CISMeF, Rouen University Hospital, France & GCSIS, LITIS EA 4108, Institute of Biomedical Research, University of Rouen, France*

^b *LERTIM, Faculte de Medecine, Universite de la Mediterranee, Marseille, France*

^c *VIDAL Company, Issy les Moulineaux, France*

Abstract. Objective: The neighbors of a document are those documents in a corpus that are most similar to it. The objective of this paper is to develop and evaluate the related resources algorithm (CISMeF-RRA) in the context of a quality-controlled health gateway on the Internet CISMeF. **Method:** CISMeF-RRA is inspired by the PubMed Related Citations Articles. CISMeF-RRA combines statistical distances with a semantic distance using MeSH terms/qualifiers. **Material:** In this feasibility study an evaluation was performed using 50 CISMeF resources randomly chosen. **Results:** Overall, 49% of the related documents were ranked as relevant. **Conclusion:** if this feasibility study is confirmed by another evaluation of more resources, CISMeF-RRA will be implemented in the CISMeF catalog

Keywords. Algorithms; Automatic Data Processing; Catalog; Medical Subject Headings

Introduction

The Internet and in particular the Web has become an extensive health information repository. In this context, several quality-controlled health gateways have been developed [1]. Quality-controlled subject gateways were defined by Koch [2] as Internet services which apply a comprehensive set of quality measures to support systematic resource discovery.

Among several quality-controlled health gateways, CISMeF ([French] acronym for Catalog and Index of French Language Health Resources on the Internet) [3] was designed to catalog and index the most important and quality-controlled sources of institutional health information in French in order to allow end-users to search them quickly and precisely (N= 36,851). CISMeF is used by Netizens and health professionals mainly from the French-speaking countries (N \approx 50,000 users per working day).

CISMeF is manually indexed by a team of four indexers, who are medical librarians. Its URLs are <http://www.chu-rouen.fr/cismef> or <http://www.cismef.org>. The Doc'CISMeF search engine has four search types: Simple, Advanced, Boolean, and Step by Step. In the Simple search, the end-user enters a query in a natural language in French or in English. This query is then automatically transformed by natural language processing tools [4] (e.g. phonemization, stemming) to map this query to the terminology used by CISMeF based on the MeSH thesaurus developed by the US National Library of Medicine [5]. The display of resources answering the end-user query is common to the four search types of the Doc'CISMeF search engine. Then, the end-user must choose the most interesting resources function of his/her context, which is most of the time much more complex than the one expressed in the query.

From one resource, it is important to obtain the nearest neighbors of a resource (or most related resources). The neighbors of a resource are those documents in the database that are the most similar to it [6].

The objective of this paper is to develop and to evaluate the algorithm "most closely related resources" in the CISMeF database. The CISMeF Related Resources Algorithm (CISMeF-RRR) is derived from the original idea of the work performed by Kim *et al.* [7]. The Related Citations Articles (NLM-RCA) feature is available from PubMed, which is a service of the US National Library of Medicine that mostly includes over 17 million citations from the MEDLINE bibliographic database. CISMeF-RRR was clearly inspired by NLM-RCA, but the algorithm was modified to adapt it to the more heterogeneous scope of Internet resources from the CISMeF gateway, when compared to scientific articles from the MEDLINE bibliographic database. The main difference of our approach consists in combining the statistical distance between documents as established by Kim *et al.* [7] with a semantic distance using the MeSH terms/qualifiers and the CISMeF resources type (RT).

1. Methods

1.1. CISMeF Terminology

The CISMeF terminology is exploited for several tasks: manually performed resource indexing, automatically performed resource categorization, visualization and navigation through the concept hierarchies in a CISMeF Terminology Server ([URL: http://www.chu-rouen.fr/terminologiecismef/](http://www.chu-rouen.fr/terminologiecismef/)) and information retrieval using the Doc'CISMeF search engine. CISMeF uses two standard tools for organizing information: the MeSH thesaurus and several metadata element sets, in particular the Dublin Core metadata set ([URL: http://www.dublincore.org/](http://www.dublincore.org/)) [10]. The MeSH terms (24,357 in 2007) are organized into hierarchies going from the most general at the top of the hierarchy to the most specific at the bottom of the hierarchy. The "is-a" and the "part-of" relations between concepts are extracted from the MeSH files to define the subsumption relationships in the CISMeF terms hierarchy.

However, the MeSH thesaurus was originally intended to index scientific articles for the Index Medicus and for the MEDLINE database. In order to customize it for the broader field of health Internet resources, we developed several enhancements [3] to the MeSH thesaurus, with the introduction of two new concepts, metaterms (MT) and resource types respectively.

A metaterm is a medical specialty or a biological science (e.g. cardiology, bacteriology), which has semantic links with one or more MeSH terms, subheadings and RTs. CISMef resource types are an extension of the publication types of MEDLINE. As defined by the Dublin Core Metadata Initiative ([URL: http://www.dublincore.org/documents/dcmi-terms/](http://www.dublincore.org/documents/dcmi-terms/)) [10], a CISMef RT (N=278) is used to categorize the kind of the content of a resource. MeSH <term/subheading> pairs describe the topic of the resource. For example, in the case of a clinical guideline about carbon monoxide intoxication, 'carbon monoxide poisoning' is the MeSH term and 'clinical guidelines' is the resource type. The RT controlled list is available at the following URL: <http://www.chu-rouen.fr/documed/typeeng.html>. The RT list has been manually built and maintained by the CISMef team since 1997.

Major Topics exist in the MEDLINE database and the CISMef catalogue for terms and qualifiers. A term is said to be "major" if the concept it represents is discussed throughout the whole document, or on the contrary "minor" if it is referred to only in a few paragraphs. Major terms are marked in MEDLINE and CISMef by a star. In CISMef, Major Topics are extended to resource types and metaterms. This task is manually performed by the CISMef medical librarians for resource types. It is automatically performed for metaterms: a metaterm is "major" for a CISMef resource if and only if at least one term, qualifier or resource type semantically linked to this metaterm is major for the same CISMef resource (otherwise, the metaterm is minor).

1.2. Similarity calculation between documents

As mentioned by Kim *et al.* [7], the similarity between documents is measured by the words they have in common, with some adjustment for document lengths. In our work, the criteria allowing similarity calculation between documents are based on the description and indexing by the CISMef medical librarians. There are four criteria as follows: Title of the document, Abstract, MeSH terms (or pairs MeSH term/subheading) and CISMef resource types. These four criteria belong to the Dublin Core metadata set [11] and comprise the overall representation of a document.

The concept of document and its representation play a fundamental part in the step of an effective computation of inter-document similarity as well as on the treatment level or on the relevance level. However, the most used representation is the vectorial representation in which a document is represented by a t-dimensional vector, where t is the total number of terms in the document database. The inter-document comparison can then be performed by a cosine measure of these two documents vectors [12]. Two steps may be used to reduce the space dimension: elimination of stopwords¹ and stemming² to reduce the grammatical variations of words to a possible root word.

Having obtained the set of terms that represents every documents, the next step is to assign a numerical weight for every stemmed word. Thus, each word will be balanced with a TF-IDF weight [12], which is computed as being the multiple of the frequency of a term in a document (TF) by the inverse weight of the frequency of the document in the collection (IDF). In this way a frequent term occurring in a small number of documents will have a greatest weight. In addition to the vectorial distance, three heuristic weightings were defined by the CISMef team: 1) in order to give an additional weight to the words in the title vs. the words in the abstract (7 and 1

¹Words with very low discrimination values in the retrieval process

²We use a stemming strategy developed in CISMef

respectively), 2) to give additional weights to major MeSH terms and major CISMef RT vs. minor MeSH terms and minor CISMef RT (7 and 3 respectively), 3) to give respective weightings to the four MeSH relations: Hierarchy, See Also, Pharmacological Action, Do Not Confuse (1, 0.1, 0.1, -0.1 respectively) reflecting their respective importance in computing the overall semantic distance.

1.3. Semantic distance

In this work, the similarity between documents also has a semantic dimension in addition to the syntactic dimension previously defined. A word-by-word distance can be defined between the MeSH terms and the MeSH subheadings. The MeSH hierarchical relation is defined as the traditional relation that exists between the concepts in a tree structure. The distance in this relation will be computed in particular by being based on the taxonomic links “is-a”, and “part-of”: the more distant in the hierarchy the two terms are, the larger the distance. There is no computation of distance for the three other relations, because for each relation there is a list of word pairs (in the relation) and they will be given a score reflecting the weight of the section 1.2. For example for the relation "Do not confuse" the two MeSH terms "sunstroke" and "heat stroke" are in connection and a score of “-0.1” will be given according to this relation.

Thus, the global semantic similarity takes into account not only the hierarchical relation (“is-a”, “part-of”) of both the MeSH thesaurus and the CISMef resource types thesaurus but also the three other relations of the MeSH thesaurus. In this semantic distance computation, we are taking into account the subheading affiliation to a MeSH term, and the RT affiliation to a MeSH term (or a MeSH term/subheading pair) [8].

Contrary to NLM-RCA [7] the CISMef-RRA takes into account Major/Minor indexing for MeSH terms, MeSH subheadings, and CISMef resource types. For the hierarchical relation the score is computed according to the more information that two terms share in the MeSH tree structure. We have chosen the Lin's similarity [9] to compute this information, already used to compute semantic distance [13].

Given two terms m_i and m_j , the Lin similarity between them is defined as:

$$sim(m_i, m_j) = \frac{2 \times \max_{m \in S(m_i, m_j)} [\log(p(m))]}{\log(p(m_i)) + \log(p(m_j))} \quad (1)$$

Where $S(m_i ; m_j)$ is the set of the ancestor terms shared by both m_i and m_j , max represents the maximum operator and $p(m)$ is the probability of finding m or any descendants in a reference corpus. It generates normalized similarity values between 0 and 1. Because Lin's similarity model relies on information content, when one term is the parent of another, their similarity is low when the parent term is placed high in the hierarchy. Conversely, it is high when the parent term is low in the hierarchy. Thus, the total similarity between the MeSH terms of two documents I and J will be measured by applying an average of the distances obtained between all their MT according to the four relations:

$$Sim(D_i, D_j) = \frac{\sum sim(a, b)}{card(D_i) \times card(D_j)} \quad (2)$$

$\forall a \in D_i$ and $\forall b \in D_j$, where D_i and D_j are set of MeSH terms of documents

I, J respectively.

Finally, the total similarity between documents will be a combination of two measurements of similarity (syntactic and semantic).

1.4. Evaluation

In order to test our algorithm we extracted from the CISMeF corpus a randomly-chosen sample of 50 resources and we run two distance algorithms (CISMeF-RRA and NLM-RCA) on this sample as a feasibility study. A manual evaluation was carried out *a posteriori* by an expert medical librarian of the CISMeF team. (CL) She quantified the number of relevant results according to a qualitative Likert scale of 5 levels, her opinion being regarded as the reference (gold standard). The evaluation was performed in two steps: Step 1: For each of the 50 resources, all the resources classified by the algorithm as "related resources" were rated by the medical librarian. Step 2: for each of the 50 resources, only the top 3 resources were rated.

2. Results

The results of the two-step evaluation are presented in Table 1. For CISMeF-RRA, overall 49% of the related resources relevant were ranked as relevant (Good or Very Good) whereas 30% of them do not reach the average (Very Bad one or Bad). In the second step of the evaluation the resources considered as the nearest (first position) were ranked relevant (Very Good or Good) in 68% of the cases, while the resources in the third position were ranked relevant in 58% of these cases.

Table 1. Step 1 and Step 2 of the evaluation

Results by position	Step1				Step2											
	All				1				2				3			
	CISMeF RRA		NLM RCA		CISMeF RRA		NLM RCA		CISMeF RRA		NLM RCA		CISMeF RRA		NLM RCA	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Very Good	33	19	17	15	18	36	16	33	14	28	13	25	11	22	12	21
Good	52	30	20	18	16	32	11	22	12	24	09	17	18	36	12	21
Good	32	18	7	06	05	10	03	6	12	24	09	17	08	16	05	08
Average	28	16	28	25	06	12	08	16	07	14	10	19	08	16	12	21
Bad	24	14	38	34	05	10	10	20	05	10	11	21	05	10	15	26
Very Bad																

3. Discussion

In this feasibility study, the CISMeF-RRA gave satisfactory results as overall 49% of the related documents were rated as "very good" or "good" vs. 37% for the NLM RCA. This feasibility study is based on a relatively small sample (N= 50). It should be

followed by a more complete evaluation based on the whole manually indexed corpus (N=21,838).

When compared to the NLM-RCA, the CISMef-RRA has several differences. The CISMef-RRA computes the inter-document similarity by using two distances. One is in common with NLM-RCA, which is based on a vectorial approach. Nevertheless CISMef-RRA is based on a weighting of the terms by using the TF-IDF in opposition to the weighting derived from the Poisson model of term frequencies in NLM-RCA. These two weighting measures are based on a similar basic concept: most frequent terms in the documents will have small weights. The main innovation of the CISMef-RRA relies on the use of semantic inter-document distance based on Lin's similarity metrics for the MeSH hierarchy relation, CISMef resource types hierarchy relations, and the semantic links between MeSH terms according to the three other relations ("See Also", "Pharmacological Action", "Do Not Confuse"). Another difference between CISMef-RRA and NLM-RCA relies on Major/Minor indexing processing, Major weighting differs from Minor one in CISMef-RRA whereas the weights are similar in NLM-RCA.

In the near future, we will need to estimate in a more convincing way the various weightings that were manually assigned by the CISMef medical librarians. We also envisage to make our semantic distance algorithm more complex by implementing several relations coming from other medical terminologies, in particular SNOMED CT semantic network. We will soon benchmark the CISMef "Related Resources algorithm" vs. NLM "Related Articles algorithm" based on the overall manually indexed CISMef corpus using a blind evaluation by a medical librarian.

References

- [1] F. Abad Garcia, A. Gonzalez Teruel, P. Bayo Calduch, R. de Ramon Frias, L. Castillo Blasco, A comparative study of six European databases of medically oriented Web resources, *J Med Libr Assoc* 93(4) (2005), 467–79.
- [2] T. Koch, Quality-controlled subject gateways: definitions, typologies, empirical overview, *Subject gateways. Online Information Review* 24(1) (2000), 24–34.
- [3] M. Douyère, LF. Soualmia, A. Névéol, A. Rogozan, B. Dahamna, JP. Leroy, B. Thirion, SJ. Darmoni, Enhancing the MeSH thesaurus to retrieve French online health resources in a quality-controlled gateway. *Health Info Libr J* 21(4) (2004), 253–61.
- [4] LF. Soualmia, B. Dahamna, B. Thirion, SJ. Darmoni, Some Strategies for Health Information Searching. *Stud Health Technol Inform* 124 (2006), 595–600
- [5] SJ. Nelson, WD. Johnson, BL. Humphreys, Relationships in Med. Subject Headings, In Relationships in the org. of knowledge, CA. Bean and R. Green, eds., Kluwer Academic Publishers, (2001), 171–84.
- [6] Computation of Related Articles. [Web access]
URL:<http://www.ncbi.nlm.nih.gov/entrez/query/static/computation.html>, Last rev.: Feb 6th, 2003.
- [7] W. Kim, AR. Aronson, J. Wilbur, Automatic mesh term assignment and quality assessment. *Proc. AMIA symposium* (2001), 319–23.
- [8] SJ. Darmoni, B. Thirion, F. Ionut-Florea, A. Rogozan, C. Letord, G. Kerdelhué, JN. Dacher, Affiliation of a resource type to a MeSH term in a quality-controlled health gateway Medinfo, 12th World Congress on Health and Medical Informatics (2007), 129:407-11.
- [9] D. Lin, An information-theoretic definition of similarity. In *Proc. Int. Conf. On Machine Learning* (1998), 296–304.
- [10] M. Dekkers, S. Weibel, State of the Dublin Core Metadata Initiative, *D-Lib Magazine* 9(4) (2003).
- [11] S. Hoelzer, RK. Schweiger, H. Boettcher, J. Rieger, J. Dudeck, Indexing of Internet resources in order to improve the provision of problem-relevant med. inf. *Stud Health Technol Inform* 90 (2002), 174–7.
- [12] G. Salton, C. Buckley, Term-weighting approaches in autom. text retrieval. *IPM* 24(5) (1988), 513–23.
- [13] A. Névéol, K. Zeng, O. Bodenreider, Besides Precision & Recall: Exploring Alternative Approaches to Evaluating an Automatic Indexing Tool for MEDLINE. *Proc. AMIA Symposium* (2006), 589–93.

Association Studies on Cervical Cancer Facilitated by Inference and Semantic Technologies: The ASSIST Approach

Pericles MITKAS^{a,b,1}, Vassilis KOUTKIAS^a, Andreas SYMEONIDIS^b, Manolis FALELAKIS^{a,b}, Christos DIOU^{a,b}, Irini LEKKA^a, Anastasios DELOPOULOS^{a,b}, Theodoros AGORASTOS^a and Nicos MAGLAVERAS^a

^a *Aristotle University, Thessaloniki, GREECE*

^b *CERTH, Informatics and Telematics Institute, Thessaloniki, GREECE*

Abstract. Cervical cancer (CxCa) is currently the second leading cause of cancer-related deaths, for women between 20 and 39 years old. As infection by the human papillomavirus (HPV) is considered as the central risk factor for CxCa, current research focuses on the role of specific genetic and environmental factors in determining HPV persistence and subsequent progression of the disease. ASSIST is an EU-funded research project that aims to facilitate the design and execution of genetic association studies on CxCa in a systematic way by adopting inference and semantic technologies. Toward this goal, ASSIST provides the means for seamless integration and virtual unification of distributed and heterogeneous CxCa data repositories, and the underlying mechanisms to undertake the entire process of expressing and statistically evaluating medical hypotheses based on the collected data in order to generate medically important associations. The ultimate goal for ASSIST is to foster the biomedical research community by providing an open, integrated and collaborative framework to facilitate genetic association studies.

Keywords. Association Studies, Integration of Biomedical Information, Semantic Technologies, Cervical Cancer, Biomedical Informatics

Introduction

Cervical cancer (CxCa) is the second leading cause of cancer-related deaths after breast cancer, for women between 20 and 39 years old [1]. Infection by the human papillomavirus (HPV) is considered as the central risk factor for CxCa [2]. However, it is unlikely to be the sole cause for developing cancer. Ongoing research investigates the role of specific genetic and environmental factors in determining HPV persistence and subsequent progression of the disease [3]. In this context, genetic association studies constitute a significant scientific approach that may lead to a more comprehensive and holistic insight on the origin of complex diseases, such as CxCa [4]. Genetic association studies aim to detect association between one or more genetic

¹ Corresponding Author: Prof. Pericles Mitkas, Department of Electrical & Computer Engineering, Faculty of Engineering, Aristotle University, 54124, Thessaloniki Greece; Email: mitkas@auth.gr.

variants (e.g. polymorphisms) and a trait, which might be some quantitative characteristic, a discrete attribute, or a disease [5]. A genetic variant is genotyped in a population for which phenotypic information is available (such as disease occurrence, or a range of different trait values). If a correlation is observed between genotype and phenotype, there is an association between the variant and the disease or trait [6].

Nevertheless, association studies are most of the times inconclusive, since the datasets employed are small, usually incomplete and of poor quality. In this regard, the ASSIST EU-funded research project aims to provide researchers with an integrated environment enabling association studies among genetic characteristics, environmental agents and viral factors, which can suggest pathogenetic mechanisms that will provide new markers of risk for CxCa. Its overall objective is to offer a new technological solution that will virtually unify multiple patient record repositories (physically located at different laboratories, clinics and/or hospitals) containing both genotypic and phenotypic data, thus, enabling researchers to utilize existing patient data from several clinics and perform research in a low-cost and time-efficient way.

In this paper, the overall system architecture and the underlying design conceptualization of ASSIST is presented, emphasising also on the medical data coding and the medical rules employed towards the unification of biomedical data and the semantic inference of medical knowledge. ASSIST's functionality is illustrated via an example usage scenario, while potential extensions and remarks on its virtue conclude the paper.

1. Rationale

The number of studies elaborating on phenotype-genotype associations for common diseases is rapidly increasing; however, several studies show variation in the underlying association between genotype and outcome between the populations studied, resulting in questionable findings. It is evident that reliable association studies require large sets of patient phenotypic and genotypic data, all provided in a structured format. In addition, significant drawbacks in current association studies are considered as the lack of standardisation in data collection, the lack of a standardised overall methodology, cost (mainly for genetic tests), time consumption and man power, failure to attempt replication of results, and the incorporation of 'disposable' study groups.

Evidently, this rather limited progress in the field is mainly due to the problems in unification and utilization of data stemming from several similar yet 'isolated' studies. Clearly, methodologies for data and system interoperability, assisted by the appropriate technologies, will alleviate some of the problems mentioned above. While standardisation of phenotypic data and clinical practices seems unrealistic, semantic annotation and transformations to widely agreed classification schemes may provide workable solutions. This approach constitutes the basis for ASSIST, which aims to be used primarily by biomedical researchers for the identification of new markers of risk, diagnosis and prognosis, and possibly treatment of CxCa via the implementation of case-control association studies.

In essence, three types of usage scenarios are supported in ASSIST, i.e. multi-criteria based data retrieval from the available medical archives, design and implementation of an association study using existing patient data from different medical institutions, and assessment of patient risk for development of CxCa based on similar available patient cases. More specifically, typical examples of use are:

- What percentage of women who have been diagnosed with *LCIN* (Low grade Cervical Intraepithelial Neoplasia) are smokers?
- Find CxCa subjects with *MTHFR* (MethyleneTetraHydroFolate Reductase) data.
- What is the association of the *MTHFR C/T* polymorphism and smoking with the development of CxCa?
- What is the risk for patients with low risk *HPV* infection and *MTHFR C/C* genotype to develop *LCIN*?

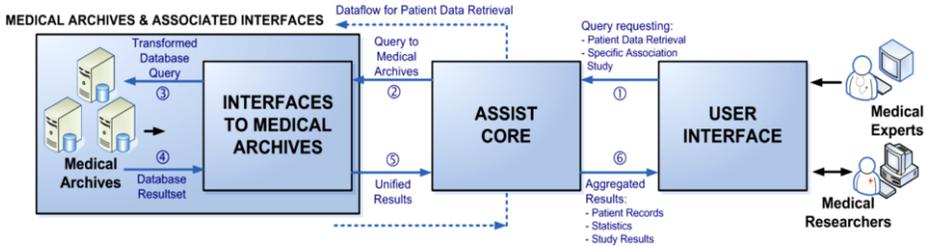


Figure 1. ASSIST framework overview and dataflow between ASSIST subsystems

2. Core Technologies and System Components

From a technical viewpoint, ASSIST aims at: i) the design of a unification mechanism of multiple patient record repositories at the syntactic and semantic level, ii) the conformance to regulations and local data access policies at each participating clinical site, iii) the specification of an automated and transparent data retrieval mechanism to support datasets definition for case-control association studies [5-6] and/or the evaluation of medical hypotheses to assess patient risk for development of CxCa, iv) the development of an inference mechanism capable of statistically evaluating medical hypotheses, and v) the design of expressive and friendly tools to perform association studies. In this regard, ASSIST follows a semantic approach [7], and resorts to medical data virtual unification and inferencing applied on real patient data. Figure 1 provides an overview of the system and the dataflow between its three subsystems, namely:

- The **Medical Archives and Associated Interfaces** subsystem. The Medical Archives that are available to ASSIST constitute research-oriented data repositories related to CxCa that ASSIST has been granted full access to. These repositories are not identical to the Hospital Information System that corresponds to each ASSIST site, but might contain a part of its hosted information, and contain only anonymised patient data, in full compliance to the legal national and EU medical research requirements and code of practice, that are generated by an anonymisation tool developed within the context of the project. ASSIST considers heterogeneous and legacy patient data repositories for integration. For each Medical Archive, the Associated Interfaces map the contained information to corresponding entities as defined in the knowledge model of ASSIST. The interfacing part between the Medical Archives and the rest of the ASSIST system ensures transparent and uniform access to patient data, confronting this way the semantic and syntactic heterogeneity of the data sources incorporated in ASSIST.
- The **ASSIST Core** subsystem. It constitutes the intermediate between the former subsystem and the User Interfaces subsystem. Based on the information provided by the Medical Archives and Associated Interfaces, the Core subsystem infers

knowledge about patients [8], and offers retrieval (query answering) and data analysis (statistical) services, supporting the definition, execution and management of association studies. It also incorporates mechanisms for handling uncertainty in data retrieval as well as query expansion and validity assessment, while it takes into account optimisation of response time with respect to the inference and retrieval procedures. Equally important, it provides the query schema to the User Interfaces subsystem and the result schema for the Associated Interfaces via its Medical Knowledge Base.

- The **User Interfaces (UI)** subsystem. It enables transparent and advanced access to the CxCa related data repositories incorporated in ASSIST. Specifically, it offers query expression as well as patient data and statistical results visualisation to the ASSIST end-users. It also incorporates modules for user profile management, i.e. a users' database, profile and session management modules, etc., as well as the means to search for previous association studies that have been performed via ASSIST. In general, the UI subsystem constitutes the front-end of the users for accessing the ASSIST services, offering a comprehensive and integrated environment for conducting association studies related research.

Figure 2 illustrates the overall architecture of ASSIST, detailed at the modules level, as well as the data exchange means between subsystems.

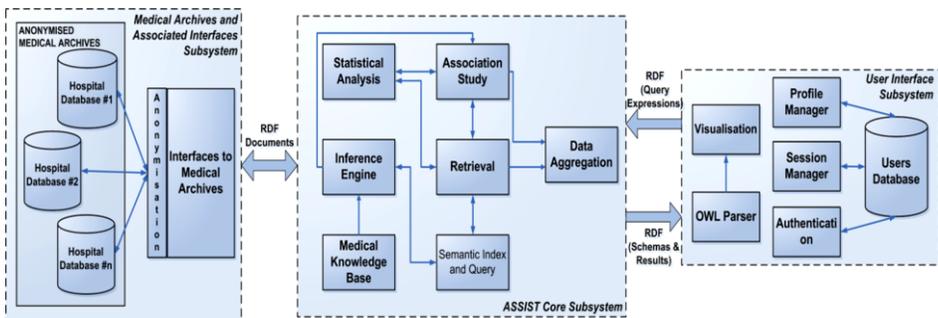


Figure 2. ASSIST overall system architecture (detailed at modules level) and data workflow

3. Medical Data Coding and Medical Rules

Currently, ASSIST integrates patient data originated from three gynaecology clinics located in Belgium, Germany and Greece, resulting in more than a thousand records. Each data repository has its own autonomous schema and captures patient information related to CxCa, according to the local clinical procedures followed. ASSIST virtually unifies these heterogeneous data sources syntactically, by defining a common data schema, and semantically via a domain ontology for CxCa that encapsulates the appropriate medical knowledge to model the disease towards inferring its severity index via the available diagnostic and therapeutic information (cytology, colposcopy, histology and histopathology), the HPV infection results, the genetic profile based on putative polymorphisms for CxCa, as well as lifestyle and personal information. As far as the examinations are concerned, a uniform coding has been elaborated for each one, so that the underlying heterogeneity in each clinic site is addressed. This uniform classification of exam results is then employed in the context of medical rules for the inference part of the system. Figure 3 illustrates the heterogeneity of example CxCa-

related biomedical data handled by ASSIST, focusing on the various coding schemes for cytology and the proposed ASSIST common coding (virtual unification part).

ASSIST considers polymorphisms of *p53*, *MTHFR*, *CYP1A1*, *CYP2E1*, *GSTM1* and *GSTT1* genes as genetic markers on the basis of their involvement in the defence against viral infections and tumor growth, as well as the number of published articles that have reported positive correlation between CxCa and genetic markers. ASSIST handles data on HPV types, as well as lifestyle and personal quality attributes, e.g. smoking, use of contraceptives, sexual activity, etc. to support the conduction of association studies.

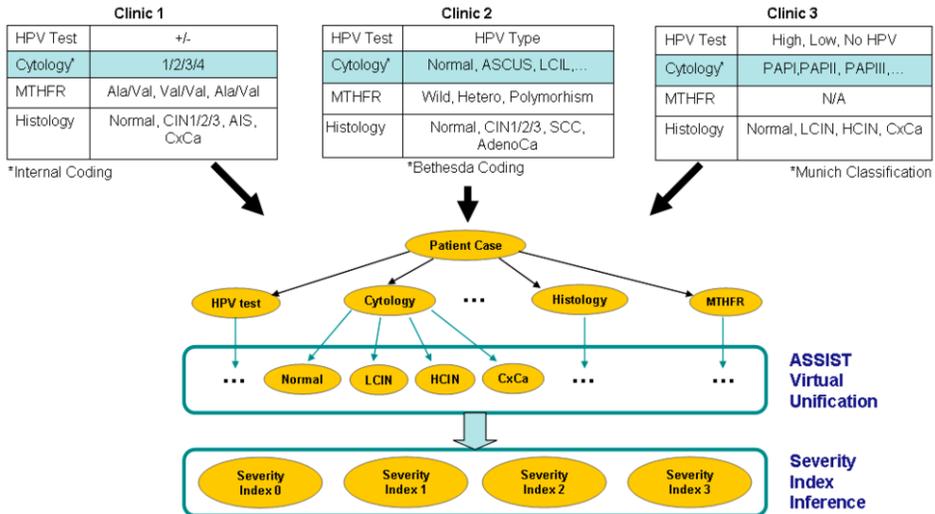


Figure 3. Heterogeneity in biomedical data related to CxCa and ASSIST virtual unification

4. Usage Scenario

From a user viewpoint, association studies involve a rather complex multi-step procedure. In these steps, several parameterisations take place, while there are also several pre/post-conditions that have to be fulfilled in order to move from one step to the subsequent one. ASSIST provides a controlled environment to ensure the validity of users' choices and offers guidance in any operation that users want to perform.

Consider a researcher wants to identify potential association of the MTHFR polymorphism with smoking in developing CxCa. Following an authentication procedure, she is granted access to the ASSIST services and selects to design an association study among the available options. For this operation, ASSIST provides her with a guidance wizard consisting of consecutive steps, i.e.: i) Data Retrieval, ii) Data Validation, iii) Genetic Description, and iv) Association Test. Specifically, she first defines the features of interest in terms of both study factors (smoking, MTHFR polymorphism) and inclusion criteria (severity of CxCa) and performs data retrieval. To ensure rich query expressiveness, a tree-like hierarchy was conceptualized and implemented in the ASSIST UI for the description of the incorporated concepts related to CxCa and patient data consisting of categories, such as *Severity of Cervical Neoplasia*, *Diagnostic Information*, *Therapeutic Intervention*, *Lifestyle*, *Personal Profile*, *Patient Profile*, *Genetic Marker* and *Specimen Availability*, and relevant

subcategories that further specify the corresponding concept. The retrieved dataset can be viewed either in detail, or through statistical measures and histograms (e.g. race and age distributions). Next step involves data validation against allele and haplotype frequencies of different populations by accessing reference polymorphism databases (e.g. HapMap). Moreover, the Hardy-Weinberg equilibrium is extracted for the case and control groups and the allele, genotype frequencies are calculated for the subject dataset in the succeeding step (Genetic Description). The results of the statistical analysis (e.g. logistic regression) are presented in the Association Tests step and the clinical significance is evaluated through the corresponding p-values.

5. Conclusion

Lately, there is an increasing research interest in genetic association studies. Several projects are under development, aiming to provide researchers with more powerful tools for investigating associations among various types of clinical and genetic data [9]. The International HapMap project (<http://www.hapmap.org/>), being one of the most ambitious ones, aims to develop a haplotype map of the human genome and provide it as a public resource that will help researchers find genes associated with human diseases, responses to drugs and environmental factors. ASSIST moves in a parallel course and, upon successful completion, its platform aspires to function as an IT tool enabling association studies linked for example with CxCa research by establishing a collaborative environment and allowing any medical group active in this area to use its facilities and/or contribute their own data/results. Following a generic design, the ASSIST system may be expanded in terms of its underlying knowledge model in order to facilitate genetic association studies for other diseases, e.g. colon cancer and cardiovascular diseases.

Acknowledgment

This work was supported in part by the IST-2004-027510 project, entitled “Association Studies Assisted by Inference and Semantic Technologies (ASSIST)”, funded by the Commission of the European Community (CEC).

References

- [1] S.H. Landis et al., Cancer statistics, 1999. *CA Cancer, J Clin* **49**(1) (1999), 8–31.
- [2] J.M.M. Walboomers et al., Human papillomavirus is a necessary cause of invasive cervical cancer worldwide, *J Pathol* **189**(1) (1999), 12–9.
- [3] T. Agorastos et al., Human papillomavirus testing for primary screening in women at low risk of developing cervical cancer. The Greek experience, *Gynaecol Oncol* **96**(3) (2005), 714–20.
- [4] J.N. Hirschhorn, K. Lohmueller, E. Byrne, K. Hirschhorn, A comprehensive review of genetic association studies, *Genet Med* **4**(2) (2002), 45–61.
- [5] H.J. Cordell, D.G. Clayton, Genetic association studies, *Lancet* **366**(9491) (2005), 1121–31.
- [6] J.N. Hirschhorn, M.J. Daly, Genome-wide association studies for common diseases and complex traits, *Nat Rev Genet* **6**(2) (2005), 95–108.
- [7] G. Antoniou, F. van Harmelen, *A Semantic Web Primer*, MIT Press, 2004.
- [8] V. Kashyap, T. Hongsrmeier, Can Semantic Web technologies enable translational medicine?, In: C.J.O. Baker CJO, K.H. Cheung (Eds.). *Semantic Web: Revolutionizing knowledge discovery in the life sciences*, Springer, 2007.
- [9] M. Dugas, Impact of integrating clinical and genetic information, *In Silico Biol* **2**(3) (2002), 383–91.

Discrete event simulation as a tool in optimization of a professional complex adaptive system

NIELSEN, Anders Lassen ^{a1}; HILWIG, Helmer ^b; KISSOON, Niranjan ^c;
TEELUCKSINGH, Surujpal ^a

Affiliations: ^a Dept. of medicine, The University of the West Indies, St Augustine, Trinidad; ^b Emergency Dept., Eric Williams Medical Science Complex, Mt Hope, Trinidad; ^c Dept. of Pediatrics, University of British Columbia, Vancouver, Canada.

Abstract. Similar urgent needs for improvement of health care systems exist in the developed and developing world. The culture and the organization of an emergency department in developing countries can best be described as a professional complex adaptive system, where each agent (employee) are ignorant of the behavior of the system as a whole; no one understands the entire system. Each agent's action is based on the state of the system at the moment (i.e. lack of medicine, unavailable laboratory investigation, lack of beds and lack of staff in certain functions). An important question is how one can improve the emergency service within the given constraints. The use of simulation signals is one new approach in studying issues amenable to improvement. Discrete event simulation was used to simulate part of the patient flow in an emergency department. A simple model was built using a prototyping approach. The simulation showed that a minor rotation among the nurses could reduce the mean number of visitors that had to be refereed to alternative flows within the hospital from 87 to 37 on a daily basis with a mean utilization of the staff between 95.8% (the nurses) and 87.4% (the doctors). We conclude that even faced with resource constraints and lack of accessible data discrete event simulation is a tool that can be used successfully to study the consequences of changes in very complex and self organizing professional complex adaptive systems.

Keywords: Discrete-event system simulation, patient flow, professional complex adaptive systems, emergency department, efficiency.

1. Introduction

Health care systems are struggling with financial limitations worldwide. The increasing demand for health care leads to the so called health care crisis [1]. The problems were highlighted by IOM [2] prompting health care systems to be more efficient [3].

The public healthcare system in Trinidad and Tobago is less than ideal. Daily headlines emphasize its' shortcomings including unacceptable long waits in the emergency departments (=ED). According to WHO 'malfunctioning health systems are

¹ Adress for correspondence: Anders.Nielsen@sta.uwi.edu

at the heart of the problem'. In addition 'countries with few resources struggle with creating infrastructures, inadequate financing, migrating doctors and nurses, and lack of basic information on health indicators' [3]. The question is how one can improve the service within the given constraints? Simulation technology is a new approach that could be used to evaluate and craft new solutions.

1.1. Why use modeling and simulation

Decision-making at first glance seems straightforward. Define the problem, diagnose its causes, design solutions, and decide how best to implement it [4]. This approach had been tried in the ED without success. So what happens when we cannot analyze us out of the problem? It leaves us with experimentation. It is the process of "doing first" trying something so that we can learn. But this approach is not reasonable in an ED [5].

When we are unable to experiment with the 'real world' could experiments in the 'virtual world' be the solution? Several papers all from the developed world indicate that simulation could be a tool [6-11]. Two elements are needed to apply simulation. The first is that a system has been identified for investigation (i.e. the ED) and the second element is that there is a problem relating to the system that needs to be corrected (i.e. waiting time and failure in treatment). The modeling activity creates an object (i.e. a model) that is used as a vehicle for experimentation – the simulation [12]. It is a discrete system because the state of the variables (i.e. patients) only changes at discrete points in time (i.e. contact with the health care staff)[13].

2. Objectives

The objectives of this paper are to evaluate the feasibility of using modeling and simulation as a management tool in an ED optimization effort in a developing country.

The goal is to identify bottlenecks from the data generated by the simulation and test possible improvements. The evaluation of the usefulness of the simulation can be broken down to: (1) evaluate the method, (2) assess the information required to build a usable model, and (3) to determine if the data collection was worthwhile [6].

3. Methods

3.1. Software

The model building and simulation software used was Arena v.10, Rockwell Software Inc.[5]. A terminating simulation over 24 hours with 500 replications was used.

3.2. The system under investigations

What kind of organization is the ED in a developing country? A holistic approach is to view it as one living organism – a complex adaptive system. As a collection of individual agents that acts in ways that are not always totally predictable, and whose actions are interconnected so that one agent's action changes the context for other

agents [14]. No agent understands everything that is going on [15]. Each agent's action is based on the state of the system at the moment (i.e. lack of medicine, unavailable laboratory investigation). It is best viewed as a professional complex adaptive system [16]. Looking for solutions that recognizes this fact will be more sustainable [17;18].

3.3. The model building

The aim of the model was to mirror the day-to-day operation. A linear entry function and a first in / first out approach were adopted since the identifications of and subsequent removal of bottlenecks was seen as a mean to maximize efficiency. A prototyping approach to model building was adapted.

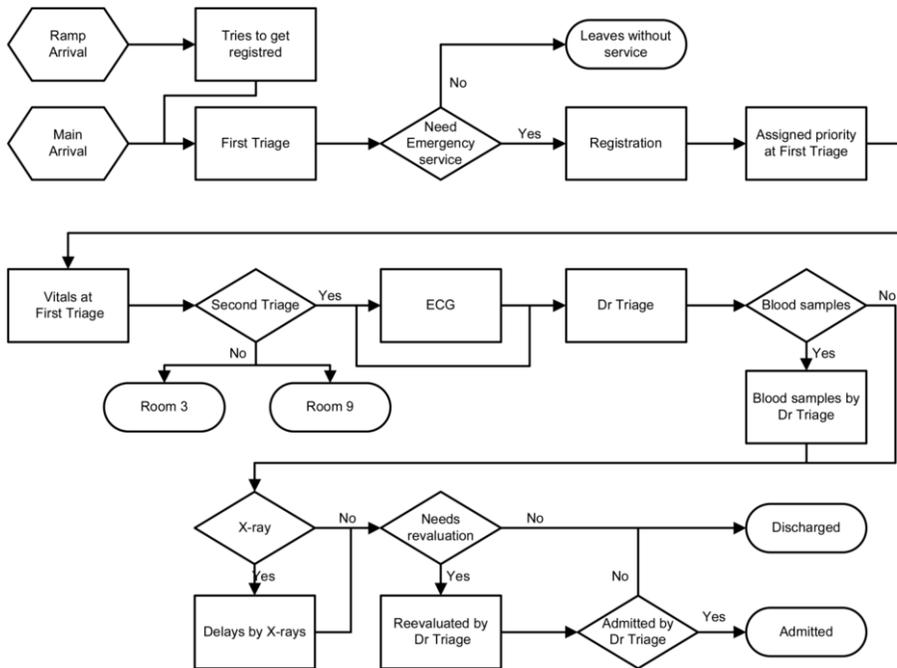


Figure 1 Emergency department flow model

The first step was interviews with senior staff to get a description of the perceived flow. Available data on admission frequencies was collected. A simple prototype model was built. Random observations of some activities were conducted to validate the model. The model was way off and raised the suspicion that far more flows existed. Hence the second step was to conduct more detailed observations. Thirty-two patients were observed (118 hours of observations). Clerical staff was used as observers to avoid interference with clinical decision and workload. Registration of individual staff performance was not done. Nine principal patient-flows were identified. An improved model was constructed. When the simulation was run it indicated a major bottleneck in the flow concerning the intake and triage. A total of 20 observed patients belonged to this flow. The remaining 12 patients were insufficient to identify the remaining eight flows type.

The third step was to build a more detailed model of the triage process (see figure 1). Based on the times observed a minimum, average and maximum time spent in each station was calculated and applied to a triangular delay type. Each station was assigned only one resource (a clerk, a triage nurse, an ECG nurse or a Doctor).

4. Results

Simulations of the existing flow with a mean of 174 (range 133-230) patients seeking emergency care resulting in accumulated average staff contact time of 36 (range 18-54) minutes. Only half of the patients were processed by the existing flow.

Simulation Flow	Fully processed N	Waiting time Hours	Utilization %			
			1 st triage	Clerk	Dr	ECG
Existing	87 (62-115)	9.2 (2.4 -17)	99.5	73.9	44.3	6.9
Improved	147 (129-135)	5.0 (0.6-13)	95.8	88.7	87.4	89.7

Table 1 Results: Simulation of existing and improved flow

One bottleneck originating from the activity ‘First Triage’ was identified. The nurse taking ‘ECG’ was only active in 7% of her time. What would happen if she could assist in taking vital signs? She was therefore exchanged with the triage nurse as the resource doing ‘vitals’. Now the simulation showed nearly a doubling of the numbers of patient processed. Total waiting time decreased and staff-utilization became more efficient.

5. Discussion

5.1. Evaluation of the method

We found like Vissers[11] that the degree of process thinking in the organization has consequences for the way the models are developed. The concept of processes is not widely shared in the observed organization. If one decides on a too detailed level, the model may be too complex for the organization to understand. If one on the other hand chooses a too high level of aggregation the information generated may not be very useful and health care practitioners may not accept the model’s description of reality. A profound insight into the dynamics of the area being modeled is therefore important requirements for developing a successful model.

We found that a simulation can indicate where changes in staffing could lead to improvement by leveling the use of resources, as is the experience from the developed world [9]. We conclude as Coats [6] does: that even thou the model were not an accurate representation of patient flow the simulation was still useful. An entry function that more closely matched the patient arrival pattern would increase the value of the simulation as a decision tool for staff planning. We found as Hung et al [9] that one could effectively change the model and easily simulate its effects on patient flow.

The conclusion is that the model building and simulation process is valuable in that it can illuminate bottlenecks and help finding solutions to solve the problem. But it has

to be emphasized that it was the whole process of identifying the flows that added most value for the organization, since several identified flows were not officially recognized.

5.2. Assess the information required to build a usable model

Harper et al.[7] argues that classification of patients should be included in a detailed model just, as detailed incorporation of recourses use should be included. However in the present case it was not practical to obtain such information. By having practitioners building the model it turned out that the initial guesses on time functions and distribution of each process were close to the ones actually observed. The prototyping approach made it feasible to quickly build meaningful models and incorporate updated knowledge as soon as it was available.

However the time consuming analysis of the activities in the organization is needed. No single expert fully grasp, and no set of documents fully captures, the subtle ways in which individual components are interwoven with one another. One has to look at the ongoing activity itself [19;20]. The employees' descriptions of their jobs correspond more or less to the formal procedure of the job manual. But when observed it is often discovered that the employees are not following those procedures at all. The employees rely on a rich variety of informal practice - not written in any manual - but crucial to getting the work done [21] but it is not necessarily efficient.

5.3. Determine if it was worthwhile to undertake the data collection needed to build a detailed model.

It was worthwhile to undertake the data collection needed to build the model. Nine performance indicators for ED[22] can be collected from the present model. That is five time definitions: Arrival-, MD contact-, decision to admit-, conversion- and discharge-time. Two time intervals: Door to doctor turnaround and discharged patients. Two defined elements of emergency: Numbers of ECGs and simple imaging procedures done. Our conclusion is that discrete event simulation can be applied to any ED. The collection of additional data and the development of more sophisticated models seem worthwhile [6].

5.4. Is it worthwhile in a developing country?

We believe so. It has been argued that the model of integrated patient pathways is a more comprehensive concept for healthcare institutions [23]. It will have to incorporate evidence-based medicine. Certainly that would require well-defined patient flow to be clinically adequate and simulation is one obvious tool to aid this development. It can be argued that potential tools have to synergistically combine people, organizational processes and technology to enable a holistic view. Simulation is one such tool.

Vissers [11] argue that modeling based health care management ought to become just as popular as evidence-based medicine. Young [24] argues that because healthcare systems around the world are undergoing redesign and refocusing on patients, there is a strategic role for modeling and simulation to play. The creation of strategic scenarios that work according to process philosophies - as used in manufacturing - could help to deliver high quality care to millions of people.

As a secondary effect of the process a picture of the real flow emerged not only the one described in the official diagrams as discussed previously but also critical aspect of

the prevailing culture surfaced. We conclude that a successful application of simulation methodology in a Third World environment was achievable and that further potential exist.

Acknowledgements: This research would not have been possible without the valuable input, time and effort afforded by many dedicated staff from within the Eric Williams Science Complex. We acknowledge the help and guidance given by our colleagues, Madaniyo Mutabazi and Jason William at the department of Engineering at the University of the West Indies, St Augustine W.I. and Craig O'Neill from the University of British Columbia, Canada.

References

- [1] Goldsmith JC. The new health-cost crisis. *Harv Bus Rev* 2001; 79(10):20-21.
- [2] Kohn LT, Corrigan JM, Donaldson MS. *To err is human: building a safer health system*. Washington, D.C.: National Academy Press; 2000.
- [3] WHO. *World Report on Knowledge for Better Health - Strengthening Health Systems*. 1-146. 2004. Geneva, Switzerland, World Health Organization. Report.
- [4] Mintzberg H, Westley F. *Decision Making: It's not What You Think*. *MIT Sloan Management Review* 2001; 42(3):89-93.
- [5] Kelton WD, Sadowski RP, Sturrock DT. *Simulation with Arena*. 4 ed. New York: McGraw-Hill; 2007.
- [6] Coats TJ, Michalis S. Mathematical modelling of patient flow through an accident and emergency department. *Emerg Med J* 2001; 18(3):190-192.
- [7] Harper PR. A Framework for Operational Modelling of Hospital Resources. *Health Care Management Science* 2002; 5(3):165-173.
- [8] Huang X-M. Decision making support in reshaping hospital medical services. *Health Care Management Science* 1998; 1(2):165-173.
- [9] Hung GR, Whitehouse SR, O'Neill C, Gray AP, Kisson NM. Computer Modeling of Patient Flow in a Pediatric Emergency Department Using Discrete Event Simulation. *Pediatric Emergency Care* 2007; 23(1):5-10.
- [10] Sinreich D, Jabali O. Staggered work shifts: A way to downsize and restructure an emergency department workforce yet maintain current operational performance. *Health Care Management Science* 2007; 10(3):293-308.
- [11] Vissers JMH. Health care management modelling: a process perspective. *Health Care Management Science* 1998; 1(2):77-85.
- [12] Birta LG, Arbez G. *Modelling and Simulation - Exploring Dynamic System Behaviour*. 1 ed. London: Springer-Verlag; 2007.
- [13] Banks J, Carson II JS, Nelson BL, Nicol DM. *Discrete-event system simulation*. 4 ed. Pearson Education Inc; 2005.
- [14] Plsek PE, Greenhalgh T. Complexity science: The challenge of complexity in health care. *BMJ* 2001; 323(7313):625-628.
- [15] Cilliers P. *Complexity & Postmodernism*. New York: The Free press; 1998.
- [16] Anderson RA. Managing health care organizations: Where professionalism meets complexity science. *Health Care Manage Rev* 2000; 25(1):83-92.
- [17] Headrick LA. *Learning to Improve Complex Systems of Care*. Prepared for Council on graduate Medical Education and national Advisory Council on nurse Education and practice joint meeting Washington D C, 2000 [cited 2005 Dec. 25]; 75-88
- [18] Nolan TW, Resar R, Haraden C, Griffin FA. *Improving the Reliability of Health Care*. Gordon AB, editor. 4, 1-17. 2004. Cambridge, Massachusetts, Institute for healthcare Improvement. *Inovation series* 2004. 21-7-2005. Report.
- [19] Hirschhorn L. *Campaigning for Change*. *Harv Bus Rev* 2002; 80(7):98-104.
- [20] Szulanski G, Winter S. *Getting It Right the Second Time*. *Harv Bus Rev* 2002; 80(1):62-69.
- [21] Brown JS. *Research that reinvents the corporation*. *Harvard Business Review on Knowledge Management*. Boston: Harvard Business School Press; 1998. 153-159.
- [22] Welch S, Augustine J, Camargo CA, Jr., Reese C. *Emergency Department Performance Measures and Benchmarking Summit*. *Acad Emerg Med* 2006; 13(10):1074-1080.
- [23] Bali K R, Dwivedi A, Naguib R. *Issues in Clinical knowledge Management: - Revisiting Healthcare management*. In: Bali R, editor. *Clinical Knowledge Management - Oportunities and Challenges*. Coventry UK: Knowledge management for Healthcare Group, Biocore, Coventry Univ.; 2005. 1-10.
- [24] Young T. *An Agenda for Healthcare and Information Simulation*. *Health Care Management Science* 2005; 8(3):189-196.

An Artificial Neural Network Derived Trauma Outcome Prediction Score as an Aid to Triage for Non-Clinicians

Adrian PEARL, Raphael BAR-OR and David BAR-OR

Swedish Medical Center and St Anthonys Central Hospital Trauma Research Lab.
Englewood, Colorado, USA

Abstract: In mass casualty events Emergency Medical Service Providers (EMS) choose treatment at Scene or a “*scoop and run*” approach. The latter requires clinically trained personnel at the reception site to triage patients. Current methodology based on Revised Trauma Score (tRTS) requires use of Glasgow Coma Scale, a method reliant on experience and clinical knowledge. This makes the system subjective and often inadequate for non-clinicians. This project attempts to develop a simplified outcome prediction score using an artificial neural network for use by a non-clinically trained EMS to aid triage. The project uses National Trauma Data Bank, Version 6.1. Tiberius Data Mining Software created Neural Network models. Variables considered were values that could easily be obtained during an event. Binary values were used for low SBP and low Respiratory Rate, coded using the RTS scoring table as a basis, and age indicators. A modified motor component of Glasgow Coma Score was created to negate the need for clinical knowledge. Best performing models, identified by Gini coefficient and ability to predict mortality, were with 8 and 10 neurons. On mortality prediction all even numbers of hidden neurons have similar performances. Training sets were compared to test sets, and found to be identical in Gini coefficient and performance. Models performed well in predicting mortality compared to standard outcome predictors. Possible additional variables such as gender or ethnicity might improve the Neural Network predictive ability. Pulse appears an essential variable not recorded by the NTDB.

Keywords: Artificial Neural Networks, mortality prediction, triage

Introduction

When there are multiple casualties in an incident, Emergency Service teams have to choose between conflicting protocols. They either provide treatment at the scene, performing *triage* and sending the serious casualties first, or operate “*scoop and run*” in which they pick-up casualties regardless of injury detail and triage as they arrive.

Triage principles whether performed in the field or at the receiving facility remain constant. Its role is to classify patients according to their medical need. With Field triage not only is it important to assess the actual severity of the injury, but location and distance to the local healthcare facilities must be considered. Additionally the resources at the disposal of the facility may affect triage decisions.

Following a traumatic injury there is a trimodal pattern to the deaths that occur¹. Immediate death occurs within minutes of injury, often due to vascular or neurological damage. Treatment rarely improves outcome. The second peak occurs during the “golden hour”. Death usually occurs due to intra-cranial haematoma, or major thoracic or abdominal injury²

Since time is critical, field triage requires a simple method to sort the injured. Standard triage is based on the Revised Trauma Score (RTS), using a simplified coded version - tRTS. It is determined by application of clinical experience to assess Glasgow Coma Score³ (GCS) and hard to apply for non-clinicians. Reliance on GCS, as has previously been argued⁴, creates a subjective system, dependent on interpretation, in which clinicians assess patient conditions. This dependency restricts the ability to obtain high correlations between injury and outcome predictions⁵. Since only clinically trained personnel could accurately determine true GCS, precious time can be lost waiting to assess injury and triage order. The Emergency Medical Services (EMS) are thus pressured into either adopting a ‘scoop and run’ policy or to ‘treat at scene’ until evacuation orders are arranged. If a new method of triage could give the EMS the ability to predict mortality at mass casualty events, it could ease pressure on Hospital Emergency Rooms while improving the survival chances of those injured.

There are a number of systematic protocols deployed to aid in first response triage, guiding injury selection based on observation of three primary functions; respiration, pulse and mental state⁶. These protocols are known by the acronym START. The variables form the basis of a simplified trauma outcome score we are attempting to design using application of artificial neural networks (ANNs) to establish predictive models to remove GCS scores from calculations⁴ in mortality prediction.

The use of ANNs in outcome prediction has become increasingly prevalent in physiological modeling. Mathematical models constructed on the basis of organic neural systems, these ANNs are flexible systems which are increasingly used in predictive modeling due to the ability of the ANN to learn and improve. This occurs by a methodology known as feed-forward/back propagation, in which the artificial neuron adds weights according to positive or negative deviation from a training set of data from which it 'learns'.

Most models use large numbers of variables, and do not align themselves for triage application. Others are determined from low volumes and restrict the ability of the model to accurately train and test the model. Our own previous work used low volumes with three measured physiological variables plus binary variants of them. Its resultant models were not as effective at low levels of sensitivity and specificity, and conclusions postulated that by using increased volume might improve the model⁴.

1. Materials & Methods

1.1.1. Patient Population

Study data was taken from the USA National Trauma Data Bank (NTDB), version 6.1 Variables extracted from the Registry included patient demographics, Scene and ED physiological variables, and the Revised Trauma Score (ED_RTS) & Revised Probability of Survival (RPS) scores for comparison against the ANN derived model. There were no exclusion criteria.

1.1.2. Creating New Variables for analysis

Neural Networks are good for classification problems, working best with data in a binary format. New variables thought to affect survival rates were created, and coded based upon the categories of the RTS, age predictors used in TRISS –Trauma Injury Severity Score, and ability to obey simple commands. These variables will be used to test for improved performance. The following variables were created as defined below:

LowSBP – Systolic Blood Pressure less than 40

LowRR - Respiratory rate shallow (less than 10)

OCmd – Ability to obey simple commands: 0 = mGCS 1-5; 1 = mGCS=6

PedAge – Age of patient class for under 16 years: 0 = >16; 1 = ≤ 16

ThirdAge – Age of Patient greater than 55 years: 0 = ≤ 54; 1 = >54

1.2. Development of the Artificial Neural Network

This study used Tiberius software (www.philbrierley.com), operating multilayer perceptron (MLPs) methodology to create ANN model algorithms. The process consists of two steps; the *forward pass*, where predicted outputs corresponding to given inputs are evaluated, and the *backward pass*, where partial derivatives are propagated back through the network. The chain rule of differentiation gives very similar computational rules for the backward pass as the ones in the forward pass. The network weights can then be adapted using any gradient-based optimization algorithm. The whole process is iterated until the weights have converged⁷.

The neural network was designed using seven input variables (three measured, two binary calculated and two age related) and one output variable. It consists of three layers, one being a hidden layer of neurons. The numbers of neurons within the hidden layer affect the number of degrees of freedom in the optimization process, and therefore the performance of the model. Adding extra neurons increases the non-linearity of the model. Therefore a higher value can be used to extract a more complicated feature. For this project 85% of the data (1,217,125) were randomly selected as a training set, and 15% (215,899) were used as the test set. There were 5011 records excluded due to missing outcome.

1.2.1. ANN design

Initial model design was based on that previously constructed for analysis of SMC data for eight input variables, using the respiratory rate, systolic blood pressure and mGCS components as recorded at Scene. Age variables were encoded as outlined above. Pulse was missing from the NTDB dataset, so only seven input variables were used for this model. Initial design started with 5 neurons in the hidden layer, and extra neurons were added to identify the best model as identified by the Gini co-efficient, and the predictive performance of the model. The Gini coefficient is a measure of equality, and can be employed as a means of comparison between ANN models⁸.

Each model design was allowed 100,000 epochs minimum in which to identify the best training algorithm before adding an extra neuron to the hidden layer. The training sets were compared in each design case to their test set, and found to be identical in Gini co-efficient and performance.

1.2.2. Statistical Analysis of Performance by Discrimination (ROC curve analysis)

Discrimination is the ability of the model to separate the population into two groups. In this instance we discriminate between those who live or die. Receiver Operator Curves (ROC) are independent of outcome prevalence, and are a useful tool in the performance evaluation for separating two populations. An ROC plot is the graph of all observed (1-specificity, sensitivity) pairs. Each point on this empirical plot can be represented by a 2x2 contingency table. Two different tests on the same patient can then be compared.⁹

1.2.3. ANN Model Analysis

Analysis of model design was by comparison of the performance in mortality prediction and survival prediction. The coefficients calculated by the ANNs were then used to calculate the probability of survival according to the following equation¹⁰

$$P_{\text{survival}} = 1 - 1 / (1 + e^b)$$

where b is calculated from:

$$b = b1(\text{lowRR}) + b2(\text{RR}) + b3(\text{SBP}) + b4(\text{lowsbp}) + b5(\text{OCmd}) + b6(\text{3rdAge}) + b7(\text{PedAge})$$

Having determined the probability of survival (mortality prediction) the values, identified as PDT, were then plotted in an ROC graph for comparison by model and with standard trauma outcome measurement scores.

2. Results

Initial design started with 5 neurons in the hidden layer. The best performing models were identified by their Gini co-efficient and ability to correctly predict mortality. The training sets were compared in each design case to their test set, and found to be identical in Gini co-efficient and performance. Results are shown in Tables 1 and 2.

Table 1: Comparison of Gini Coefficient & RMS error in ANN models

Nos Neurons	5	6	7	8	9	10	11	12
Gini	0.6204	0.6228	0.6284	0.6153	0.6187	0.6180	0.6228	0.6192
RMS Error	0.1997	0.1997	0.1999	0.1999	0.1997	0.1998	0.1997	0.1998

The best performing models, as identified by their Gini co-efficient and ability to correctly predict mortality were constructed with 8 and 10 neurons, though on mortality prediction all the models with an even number of hidden neurons have similar performances. None of the models performed as well in predicting mortality when compared to using the model that included values for patient pulse⁴.

Table 2: Comparison of Predictive performance in ANN models

Nos Neurons	5	6	7	8	9	10	11	12
% Alive	72.4	85.3	71.5	85.0	72.7	83.0	73.4	85.7
% Died	78.2	65.7	79.8	66.1	77.7	69.4	77.5	65.2

2.1.1. Statistical Analysis of Performance by Discrimination (ROC curve analysis)

For each of the best performing models (pdt8 & pdt10, 8 and 10 neurons respectively) probability statistics were calculated and then the ROC graph plotted, with the ED Revised Trauma Score and the TRISS Probability of Survival included as comparisons. In addition a five neuron model (pdt5) is included since previous research indicated that a five neuron model worked well. In comparisons it is clear that the models work as well as current outcome predictors in mortality prediction.

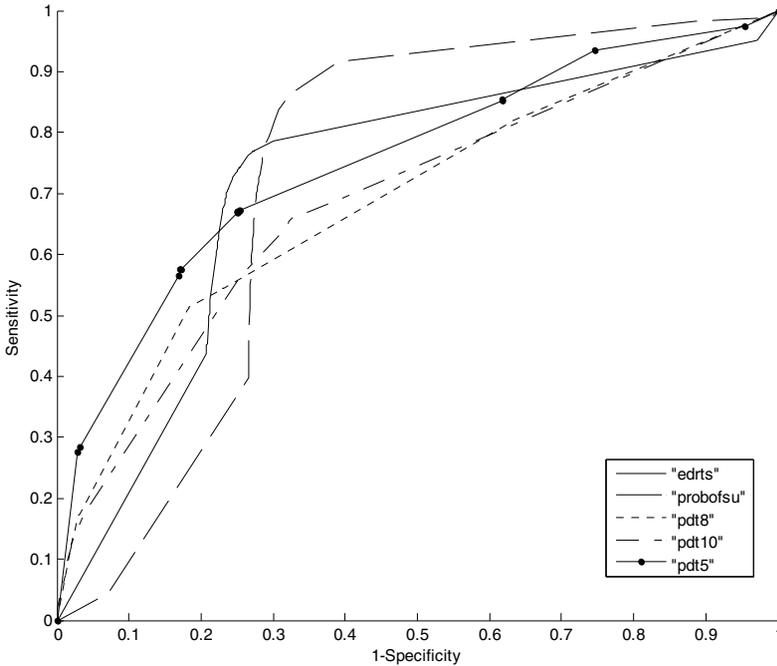


Figure 1. Receiver Operator Curve for artificial neural network models and comparison mortality predictors

3. Discussion

Using the greater volume of data improves the ability of the ANN to train, and therefore apply its predictive equation to a test set. However, when compared with the performance of the original model determined from the SMC registry data, with a Gini co-efficient of 0.493 and RMS error of 0.186, we can see that predictive accuracy is lost when excluding the patient pulse from the equation.

The original intent of the project was to create a model that could easily be applied by non-clinicians to more accurately predict mortality in a triage situation without requiring lots of variables to be taken into consideration. This analysis shows that while increased numbers improves the models ability to train, decreased variables in the model reduces the accuracy.

Using the NTDB dataset could be detrimental, as large volume of data may cause the ANN to loose important trends that could aid in mortality prediction. It may be that

the size of the population cohort used to create an artificial neural network has an optimum level.

4. Conclusion

Previous ANN-based studies conducted used population-based registries, with large numbers of inputs^{11, 12}. This study concentrated on a smaller number of factors, previously considered to give good correlation in survival prediction, to create an efficient simple method for 'field triage'. The method was designed to require no application of clinical knowledge, and relate directly to physiological characteristics that could be measured by emergency service personnel.

Compared to the current outcome predictors, these models performed well in mortality predicting. However when compared to the model which included the patient pulse, as previously reported⁴, none of the models performed as well. The pulse appears to be an essential variable in this model and needs to be recorded in the NTDB.

Further work using multi-center data to provide sufficient sample sizes that include pulse should be performed to try and improve the methodology. Possible additional variables such as gender or ethnicity might further improve the Neural Network predictive ability. Additional studies may need to be conducted to identify if population cohort size affects the ability of Neural Networks to predict mortality, and that systems designed on too small as well as too large a population dataset mask important trends.

References

- [1] Meislin H, Criss EA, Judkins D, Berger R, Conroy C, Parks B, Spaite DW, Valenzuela TD. Fatal trauma: the modal distribution of time to death is a function of patient demographics and regional resources. *J Trauma* 1997;43(3):433-40.
- [2] Surgical Tutor Online Resources – Assessment of the Injured Patient: <http://www.surgical-tutor.org.uk/default-home.htm?core/trauma/atls.htm~right>
- [3] Teasdale G, Jennett B. Assessment of Coma and impaired consciousness. *Lancet* 1974;2(7872):81-4
- [4] Pearl A, Caspi R, Bar-On D. Artificial neural network versus subjective scoring in predicting mortality in trauma patients. *Stud Health Technol Inform*. 2006;124:1019-24
- [5] Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, Flanagan ME. A revision of the Trauma Score. *J Trauma* 1989;29(5):623-9
- [6] CERT Los Angeles Simple Triage And Rapid Treatment: <http://www.cert-la.com/triage/start.htm>
- [7] Haykin, S. (1994), *Neural Networks: A Comprehensive Foundation*, NY: Macmillan
- [8] Brown M. Using Gini-style indices to evaluate the spatial patterns of health practitioners: theoretical considerations and an application based on Alberta data. *Soc Sci Med* 1994;38(9):1243-56.
- [9] Lett RR, Hanley JA and Smith JS. The Comparison of Injury Severity Instrument Performance Using Likelihood Ratio and ROC Curve Analysis. *J Trauma* 1995;38(1):142-8.
- [10] Boyd CR, Tolson MA, Copes WS. Evaluating Trauma Care: TRISS method. *J Trauma* 1987;27:370-8.
- [11] DiRusso SM, Sullivan T, Holly C, Cuff SN, Savino J. An artificial neural network as a model for prediction of survival in trauma patients. *J Trauma* 2000;49(2):212-20
- [12] Nimgaonkar A, Karnad DR, Sudarshan S, Ohno-Machado L, Kohane I. Prediction of mortality in an Indian intensive care unit. *Intensive Care Med* 2004;30(2):248-53

Case-Based Reasoning to Explain Medical Model Exceptions

Rainer SCHMIDT^a and Olga VOROBIEVA^{a,b}

^a *Institute for Medical Informatics and Biometry, University of Rostock, Germany*

^b *Sechenov Institute of Evolutionary Physiology and Biochemistry,
St. Petersburg, Russia*

Abstract. In medicine many exceptions occur. In medical practise and in knowledge-based systems too, it is necessary to consider them and to deal with them appropriately. In medical studies and in research exceptions shall be explained. We present a system that helps to explain cases that do not fit into a theoretical hypothesis. Our starting points are situations where neither a well-developed theory nor reliable knowledge nor a proper case base is available. So, instead of reliable theoretical knowledge and intelligent experience, we have just some theoretical hypothesis and a set of measurements. In this paper, we propose to combine CBR with a statistical model. We use CBR to explain those cases that do not fit the model. The case base has to be set up incrementally, it contains the exceptional cases, and their explanations are the solutions, which can be used to help to explain further exceptional cases

Keywords: Knowledge-based systems, Expert systems, Clinical trials

1. Introduction

In medicine many exceptions occur. In medical practise and in knowledge-based systems too, these exceptions have to be considered and have to be dealt with appropriately. In our previous work, we demonstrated advantages of Case-Based Reasoning (CBR) in situations where a theoretically approved medical decision does not produce the desired and usually expected results [1].

In medical studies and in research exceptions shall be explained. We have developed a system, called ISOR, that deals with situations where neither a well-developed theory nor reliable knowledge nor a proper case base is available. So, instead of theoretical knowledge and intelligent experience, we now have just some theoretical hypothesis and a set of measurements. In such situations the usual question is how do measured data fit to hypotheses. To statistically confirm a hypothesis it is necessary, that the majority of cases fit the hypothesis. Statistics determines the exact quantity of necessary confirmation [2]. However, usually a few cases do not satisfy the hypothesis. We examine these cases to find out why they deviate from the hypothesis. This approach is justified by a certain mistrust of statistical models by doctors, because modeling results are usually unspecific and "average oriented" [3], which means a lack of attention to individual "imperceptible" features of concrete patients.

Our approach starts in a situation where no case-base with complete solutions is available but has to be set up incrementally. So, we must

1. Construct a model,
2. Point out the exceptions,
3. Find causes why the exceptional cases do not fit the model, and
4. Develop a case-base.

So, we combine CBR with a statistical model. The idea to combine CBR with other methods is not new. For example Care-Partner resorts to a multi-modal reasoning framework for the co-operation of CBR and Rule-based Reasoning [4]. Another way of combining hybrid rule bases with CBR is discussed by Prentzas and Hatzilgeroudis [5]. The combination of CBR and model-based reasoning is discussed in [6]. Statistical methods are used within CBR mainly for retrieval and retention (e.g. [7]).

Dialysis and fitness. Hemodialysis means stress for a patient's organism and has significant adverse effects. Fitness is the most available and a relative cheap way of support. It is meant to improve a physiological condition of a patient and to compensate negative dialysis effects. One of the intended goals of this research is to convince the patients of the positive effects of fitness and to encourage them to make efforts and to go in for sports actively. This is important because dialysis patients usually feel sick, they are physically weak, and they do not want any additional physical load [8]. At our University clinic in St. Petersburg, a specially developed complex of physiotherapy exercises including simulators, walking, swimming etc. was offered to all dialysis patients but only some of them actively participated, whereas some others participated but were not really active.

2. Incremental Development of an Explanation Model

For each patient a set of physiological parameters was measured. These parameters contain information about burned calories, maximal power achieved by the patient, oxygen uptake, oxygen pulse, lung ventilation and others. There are also biochemical parameters like hemoglobin and other laboratory measurements. More than 100 parameters were considered. The parameters are supposed to be measured four times during the first year of participating in the fitness program. There is an initial measurement followed by a next one after three months, then after six months and finally after a year. Unfortunately, since some measurements did not happen, many data are missing. Therefore the records of the patients often contain different sets of measured parameters. It is necessary to note that parameter values of dialysis patients essentially differ from those of non-dialysis patients, especially of healthy people, because dialysis interferes with the natural, physiological processes in an organism. For statistics, this means difficulties in applying statistical methods based on correlation and it limits the usage of a knowledge base developed for normal people. Inhomogeneity of observed data, many missing values, many parameters for a relatively small sample size, all this makes our data set practically impossible for usual statistical analysis.

2.1. Setting up a Model

We start with a medical problem that has to be solved based on given data. In our example it is: "Does special fitness improve the physiological condition of dialysis

patients?" More formal, we have to compare physical conditions of active and non-active patients. Patients are divided into two groups, depending on their level of activity, active patients and non-active ones. According to our assumption active patients should feel better after some months of fitness, whereas non-active ones should feel rather worse. We have to define the meaning of "feeling better" and "feeling worse" in our context. A medical expert selects appropriate factors from ISOR's menu. It contains the list of field names from the observed database. The expert selects the following main factors

- F1: O2PT - Oxygen pulse by training
- F2: MUO2T - Maximal Uptake of Oxygen by training
- F3: WorkJ – performed Work (Joules) during control training

Subsequently the "research time period" has to be determined. Initially, this period was planned to be twelve months, but after a while the patients tend to give up the fitness program. This means, the longer the time period, the more data are missing. Therefore, we had to make a compromise between time period and sample size. A period of six months was chosen.

The next question is whether the model shall be quantitative or qualitative? The observed data are mostly quantitative measurements. The selected factors are of quantitative nature too. On the other side, the goal of our research is to find out whether physical training improves or worsens the physical condition of the dialysis patients. We have to compare each patient with his/her own situation just before the start of the fitness program. The success shall not be measured in absolute values, because the health statuses of patients are very different. Thus, even a modest improvement for one patient may be as important as a great improvement of another. Therefore, we simply classify the development in two categories: "better" and "worse". The changes are assessed depending on the number of improved factors:

- Weak version of the model: at least one factor has improved
- Medium version of the model: at least two factors have improved
- Strong version of the model: all three factors have improved

The final step means to define the type of model. The easiest model is a 2x2 frequency table. Our "Better/ Worse" concept fits this simple model very well. So the 2x2 frequency table is accepted. The results are presented in table 1.

Table 1. Results of Fisher's Exact Test, performed with an interactive Web-program:

<http://www.matforsk.no/lola/fisher.htm>

Improvement mode	Patient's physical condition	Active	Non-active	Fisher Exact p
Strong	Better	28	2	< 0.0001
	Worse	22	21	
Medium	Better	40	10	< 0.005
	Worse	10	12	
Weak	Better	47	16	< 0.02
	Worse	3	6	

According to our assumption after six months of active fitness the conditions of the patients should be better. Statistical analysis shows a significant dependence between the patient's activity and improvement of their physical condition. Unfortunately, the most popular Pearson Chi-square test is not applicable here because of the small values "2" and "3" in table 1. But Fisher's exact test [3] can be used. In the three versions shown in table 1 a very strong significance can be observed. The smaller the value of p , the more significant is the dependency.

Exceptions. Though the performed Fisher test confirms the hypothesis, there are exceptions, namely active patients whose health conditions did not improve. These exceptions should be explained. Already explained exceptions build the case base. According to table 1, the stronger the model, the more exceptions can be observed.

2.2. Setting up a Case Base

The case-base is set up sequentially. That means, as soon as an exception is explained, it is incorporated into the case-base and can be used to help explaining further exceptional cases. We chose a random order for the exceptional cases. In fact, we took them in alphabetical order.

The retrieval of already explained cases is performed by keywords. The main ones are "problem code", "diagnosis", and "therapy". In the situation of explaining exceptions for dialysis patients the instantiations of these keywords are "adverse effects of dialysis" (diagnosis), "fitness" (therapy), and two problem specific codes. Besides the main ISOR keywords additional problem specific ones are used. Here the additional key is the number of worsened factors. Further keywords are optional. They are just used when the case-base becomes bigger and retrieval is not simple any longer.

However, ISOR does not only use the case-base as knowledge source but further sources are involved, namely the patient's individual base (his medical history), observed data (partly gained by dialogue with medical experts), and expert knowledge.

Expert knowledge can be used in many different ways. Firstly we use it to acquire rules, secondly it can be used to select appropriate items from the list of retrieved solutions, and to propose new solutions.

For the problem: "Why did some patients conditions become worse, although they actively participated in the fitness program?" we got obtained a set of solutions of different origin and different nature. There are three categories of solutions: additional factor, model failure, and wrong data.

The most important and most frequent solution is the influence of an additional factor. Only three main factors are obviously not enough to describe all medical cases. Unfortunately, for different patients different additional factors are important. When ISOR has discovered an additional factor as explanation for an exceptional case, the factor has to be confirmed by a medical expert before it can be accepted as a solution. One of these factors is Parathyroid Hormone (PTH). An increased PTH level sometimes can explain a worsened condition of a patient. PTH is a significant factor, but unfortunately it was measured only for some patients. Some exceptions can be explained by indirect indications. One of them is a very long time of dialysis (more than 60 months) before a patient began with the training program. Another solution is a phosphorus blood level. One patient record contained many missing data. The retrieved

solution meant high PTH, but since PTH data in the query of the patient record was missing too, an artificial case was created, which inherited all retrieval attributes of the query case while the other attributes were recorded as missing. According to the expert high phosphorus can explain the solution. We regard two types of model failures. One of them is neglected data. In fact, three of the patients did not show an improvement in the considered six month but in the following, neglected six months. So, they were wrongly classified and should really belong to the “better” category. The second type of model failure is based on the fact that the two-category model was not precise enough. Some exceptions could be explained by a tiny and not really significant change in one of the main factors.

2.3. Illustration of the Program Flow

Figure 1 shows the main dialogue of ISOR where the user at first sets up a model (steps one to four), subsequently gets the result and an analysis of the model (steps five to eight), and then attempts to find explanations for the exceptions (steps nine and ten). Finally the case base is updated (steps eleven and twelve).

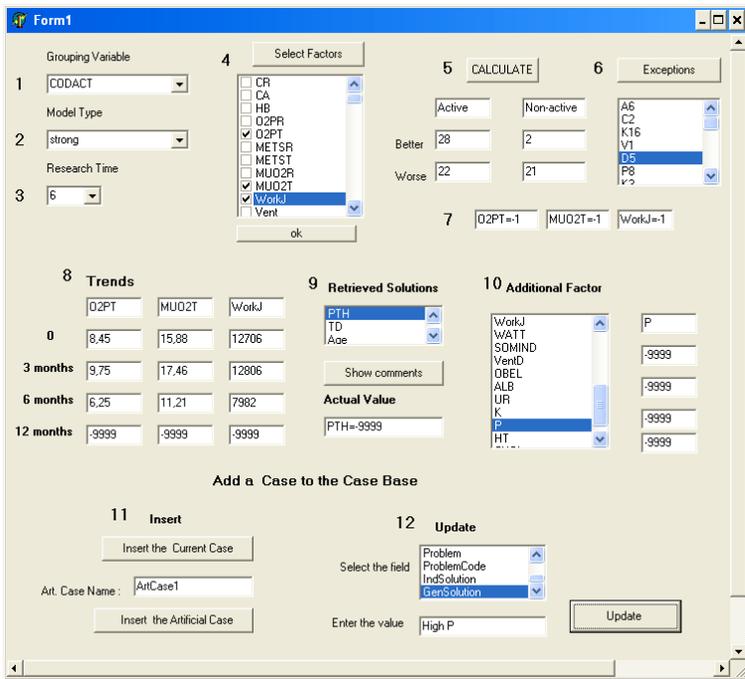


Figure 1. ISOR’s program flow

Now we explain the steps in detail. At first the user has to set up a model. To do this he has to select a grouping variable. In this example CODACT was chosen. It stands for “activity code” and means that active and none active patients are to be compared. Provided alternatives are the sex and the beginning with the fitness program (within the first year of dialysis or later). In another menu the user can define further alternatives. Furthermore, the user has to select a model type (alternatives are “strong”, “medium”,

and “weak”), the length of time that should be considered (3, 6 or 12 months), and main factors have to be selected. Here, the main factors are: O2PT, MUO2T, and WorkJ. When the user has selected these items, the program calculates the table. ISOR does not only calculate the table but additionally extracts the exceptional patients from the observed database. In the menu, the list of exceptions shows the code names of the patients. In the example patient “D5” is selected” and all further data belong to this patient. The goal is to find an explanation for the exceptional case “D5”.

In point seven of the menu it is shown that all selected factors worsened (-1), and in point eight the factor values according to different time intervals are depicted. All data for twelve months are missing (-9999).

The next step means creating an explanation for the selected patient “D5”. From the case base ISOR retrieves general solutions. The first retrieved one in this example, the PTH factor, denotes that the increased Parathyroid hormone blood level may explain the failure. Further theoretical information (e.g. normal values) about a selected item can be received by pressing the button “show comments”. The PTH value of patient “D5” is missing (-9999). From menu point ten the expert user can select further probable solutions. In the example an increased phosphorus level (P) is suggested. Unfortunately, phosphorus data are missing too. So, the expert has to find another explanation.

3. Conclusion

In this paper, we have proposed to use CBR to explain cases that do not fit a statistical model. Here we used one of the simplest models. However, it is relatively effective, because it demonstrates statistically significant dependencies, in our example between fitness activity and health improvement of dialysis patients, where the model covers about two thirds of the patients, whereas the other third can be explained by applying CBR. The presented method makes use of different sources of knowledge and information, inclusive medical experts. It seems to be a very promising method to deal with a poorly structured database, with many missing data, and with situations where cases contain different sets of attributes.

References

- [1] R. Schmidt, O. Vorobieva, ISOR: Case-based reasoning investigation of therapy inefficacy. *Knowledge-Based Systems* **19** (2006), 333-340.
- [2] M.G. Kendall, A. Stuart, *The advanced theory of statistics*. Macmillan Publishing, New York, 1979.
- [3] G.A. Hai, *Logic of diagnostic and decision making in clinical medicine*, Politeknika publishing, St. Petersburg, 2002.
- [4] I. Bichindaritz, E. Kansu, K.M. Sullivan, *Case-based Reasoning in care-partner*, Proceedings EWCBR-98, Springer, Berlin, 1998, pp. 334-345.
- [5] J. Prentzas, I. Hatzilgeroudis, *Integrating hybrid rule-based with case-based reasoning*, Proceedings ECCBR 2002, Springer, Berlin, 2002, pp. 336-349.
- [6] L. Shuguang, J. Qing, C. George, *Combining case-based and model-based reasoning: a formal specification*, Proceedings APSEC'00, 2000, pp. 416
- [7] J.M. Corchado et al.: *Maximum likelihood Hebbian learning based retrieval method for CBR systems*, Proceedings ICCBR 2003, Springer, Berlin, 2003, pp. 107-121
- [8] A.M. Davidson, J.S. Cameron, J.-P. Grünfeld J.-P. et al. (eds.), *Oxford Textbook of Nephrology*, Volume 3. Oxford University Press, Oxford, 2005.

Diagnostic Games: from Adequate Formalization of Clinical Experience to Structure Discovery

Michael A. SHIFRIN, Eva I. KASPAROVA¹
N.N.Burdenko Neurosurgical Institute, Moscow, Russia

Abstract. A method of obtaining well-founded and reproducible results in clinical decision making is presented. It is based on "diagnostic games", a procedure of elicitation and formalization of experts' knowledge and experience. The use of this procedure allows formulating decision rules in the terms of an adequate language, that are both unambiguous and clinically clear.

Key words: clinical decision making, formalization, diagnostic games, experts' knowledge

Introduction. Evidence-based medicine and clinical experience evidence

Evidence-based medicine (EBM) and computational Machine Learning (ML) approaches became very popular in the attempt to bring a scientific support to the practice of clinical medicine. Their role is considerable and undeniable. However, an important gap exists between empirical evidence and clinical practice because evidence resulting from randomized controlled study or from computational models is not directly applicable to individual patients.

Considering a typical example of a ranking system of evidence we find something like following: "*The first level*: Evidence obtained from at least one properly designed randomized controlled study <...> *The last level*: Opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees".

In this hierarchy, empirical evidence obtained from a randomized controlled study is listed first. Consequently, this "evidence" viewed as the "best" on which to make a clinical decision and supersedes clinical experience that has a lowest rank in this list. As noted by M.R.Tonelli [1], clinical experience "differs in kind, not degree, from empirical evidence and does not belong on a graded hierarchy". In accordance with this he suggests to view this kind of medical knowledge as complementary to empirical evidence. He concludes that its incorporation in EBM is "necessary to overcome the intrinsic gap noted above." But it seems clear that clinical experience or a specialist's opinion by itself is not yet what should be considered as evidence. So, what we need is

¹ M.A. Shifrin and E.I. Kasparova contributed equally to this article.

a *formalization method* that could produce a radically different type of medical evidence, namely, *evidence from clinical experience*.

Formalization of clinical experience is a sophisticated task because two antagonistic criteria must be satisfied: to formalize clinical concepts and features, and at the same time remain faithful to its deep insights and common sense. We call such formalization “adequate formalization”. Adequate formalization is decisive for the success of further mathematical processing and is the basis for the later stages of the work. In this paper we introduce an original experimental method aiming for adequate formalization of clinical experience. This method yields verifiable and reproducible results. The purpose of this paper is methodological; we will show the way to formulate and solve problems of clinical decision-making so that the results can be directly applicable to individual patients. The present approach was first developed in the collaborative works of Israel M.Gelfand and colleagues with clinicians in 1970s [2]. In this article we pursue this line of work.

1. Problems to solve

Problems we are tackling here constitute the essential part of patient oriented decision-making problems that emerge during the course of patients' treatment. Here are the typical examples of problems focusing on individual patients.

- "What is the diagnosis for the given patient?"
- "Is this drug effective for the given patient's outcome?"
- "What would the preferred treatment for the given patient be?"
- "What is the prognosis for the course of the disease for the given patient?"

All these problems are tied to clinical investigations with the aim of improving the treatment's results for a particular group of patients and of indicating ways to achieve this aim.

The analysis and formalization of a decision-making problem in a given clinical setting typically will lead to another patient-oriented problem which we will consider as a *formal decision problem* of the study. We want to stress the importance of the relation path:

aim of clinical investigation → *medical decision-making problem* → *formal decision problem of the study*.

We emphasize that clinically meaningful evaluation of results of the study should include the evaluation of improvement in the treatment in precise clinical settings. This means that the evaluation of the result should depend on to what degree the *aim of clinical investigation* has been achieved.

2. Clinical experience versus medical knowledge

Our formalization method for patient oriented clinical-decision problems extensively uses the *Diagnostic games (DG)* technique [3] and statistical analysis.

Statistical methods can be applied to a group of patients. Used alone they are often not sufficient for patient-oriented clinical problem [e.g.4]. The reasons for this are the following.

- Results obtained in terms of mathematical statistics need to be interpreted in clinical terms that may not always be obvious and unambiguous. For example,

if a reliable calculation of probability of success of a surgery intervention is not close to 1 or to 0, then it is of little use for a doctor who has to make a decision for a specific patient.

- The results obtained by formal methods depend on the data they use. Clinical data are tied with meaning, knowledge, data processing, data organization, etc. This data needs to be formalized in a way so that we keep its essential meaning in a particular clinical situation.
- Subjective information should not be ignored. Instead, more attention should be paid to the formalization methods of individual knowledge.

Specialists in machine learning and probabilistic modeling dealing with clinical applications point out the insufficiency of machine models alone and "the risks of the unwary use of data sets" [5]. Druzdzel and Diez in [5] claim that "it is necessary to resort to human experts' knowledge".

Let us give some examples from the literature of the common uses of expert knowledge.

- Expert interviewed by knowledge engineer who organizes the information elicited from the expert into a collection of rules
- Expert estimates
- Physician giving his opinion on the significance of a specific region on the ROC curve [6]
- Physician taking part in defining the graphical structure of a model (in Causal Probabilistic Models) [5,7,8,9]
- Physician describing explicitly the reasons that led him to make a certain diagnosis or recommend a certain therapy [10]

Our method for adequate formalization of clinical experience extensively uses an original technique of DGs [3], a powerful tool for clinical experience elicitation. We want to point out that the use of DGs in our method is radically different from traditional "expert knowledge" approach. In fact, our approach differs from all other proposed approaches in the following ways.

- We elicit particular types of medical knowledge - clinical experience, implicit knowledge, tacit assumptions - *i.e.* what is called "knowing-in action" [11]
- We do not just elicit knowledge from the clinician which might be ambiguous. Our actual goal is the construction of an *adequate formalized language* using elicited experience.
- Adequate language of this kind consists, roughly, of structural units and rules.
- Resulting formalized language is well founded both with respect to the logic and the actual meaning of the clinical situation considered.

3. Stages of Adequate Formalization Method

The important thing to know about Adequate Formalization (AF) method is that some of its stages are developed in parallel and some in sequential order. DGs are used in most of them. Because of size limitations we cannot describe here in detail all the stages of AF and point out their essential interrelations. We outline them below.

Formalization of a problem statement:

- A path: medical aim → medical decision problem(s) → formal decision problem of the given study
- Description of a group of patients

- Source and scope of data used by mathematicians
- Predictive concepts (elements of adequate language)
- A verification procedure for predictions

Developing of multipurpose questionnaires (principles and requirements):

- Verification of completeness
- Verification of reproducibility of data
- Formalizations of questions and answers

Elaboration of adequate language:

- Elaboration of new concepts for patients' description
- Concepts elaboration for predictions
- Developing of reduced questionnaires

Structural units discovery:

- Two levels of description: features (lower level) and structural units (upper level)
- Formal transformation rules that relate these two levels.

Construction of algorithm (decision rule) based on structural units.

Verification procedure for the algorithm:

- Verification on retrospective material
- Clinical verification (prospective study)

Medical Evaluation of the result obtained.

4. Adequate Formalization in the study of the Differential Diagnosis of Purulent Meningitis in Infants

It is well-known that the choice of an adequate treatment of a purulent meningitis depends on its etiology. The goal set for the DG carried out in the early stage of this study was defined as followed: using clinical picture of the illness the expert physician had to determine the etiology of meningitis².

Analyzing this DG it became clear that the pattern of the onset of the illness was relevant to its etiology. In the DGs that followed, the expert repeatedly came back to assessing the acuteness of the progression of illness using expressions such as "acute onset", "subacute progression of the illness", "undulating illness course" etc. However, formulating any formal criteria proved practically impossible at that point. In particular, the impediment to formalized assessment of illness onset lay in the fact that the information concerning the illness onset was communicated by patient's parents. Evidently, such information may have been incomplete and self-contradictory.

In order to find the formal criteria that indicate the illness onset a thorough application of the adequate formalization method was performed. As the first step it has proved important to understand and to express formally the notion "The first day of illness" as a formalized feature (about the two levels of formalization see [3]).

It is important to know that, in the later stages of formalization five types of illness onset were singled out and have contributed to the elaboration of adequate language concepts. These became the basis for application of the formal rules.

In the next stage four structural units have been built [2] (two of them coincided with two types of illness onset). The solution of the problem has been obtained in the

² The authors must remind that this example serves only as methodological illustration and its medical meaning will not be discussed in the present paper.

form of a relatively simple algorithm based on those structural units. The prospective verification of the algorithm and its clinical impact evaluation gave meaningful and statistically significant results.

Here are the structural units for determining the etiology of meningitis for infants during the first year of life.

1. Evident neurological symptomatology.
2. Otitis
3. Slack prodromal period with catarrhal symptoms
4. Acute prodromal period of a nasopharyngitis type

Though structural units have a simple shape some of them have complex internal structures. It is highly important that structural units

- are always expressed in a language that is understandable to physicians;
- do not require any additional interpretation at all;
- are formalized and therefore can be used in a computer algorithm.

These characteristics are essential for an adequate language.

The algorithm obtained in this study was formulated as a simple logical expression (equivalent to a Decision Tree) using structural units mentioned above. While the questionnaire elaborated in this work contained many dozens of low level features the resulting algorithm is compact and uses only 4 structural units from the adequate language vocabulary.

5. Novelty and originality of AF Method

Herein we list several essential features which define the novelty and originality of the method of adequate formalization.

- In the AF method, unlike in other approaches to formalization of experts' skills and knowledge, an elaboration of adequate description and discovery of structural units always precede construction of the algorithm. This is a radically different approach to formalization and modeling.
- The structure of the algorithm is usually rather simple. By contrast, the concepts of adequate language have often a complicated and non-trivial structure with respect to initial features.
- The AF method allows a drastic reduction of the feature space dimension. Only then, the algorithm is applied to structural units.
- The result of the algorithm coming from the AF method does not require any interpretation. The algorithm is formulated in terms that could be applied to the individual patient.
- The AF method is applicable to small size samples when the initial dataset is insufficient for statistical analysis.

6. Limitations

As with any scientific method it is important to indicate explicitly its limitations. The two most significant are as follows.

- The use of AF method is effective in those medical decision problems where sufficient clinical experience is available.
- Results obtained by the AF method are valid only in given clinical settings.

Conclusion

At a first glance, it might look like the results of AF method are not so helpful for a high level expert since in most cases, the expert makes right decisions. "However people often forget what great losses are borne by doctors and patients to attain this level of expertise" [12, p 6]. Such is the opinion of Professor AL Syrkin (Sechenov Moscow Medical Academy), a high level specialist in cardiology who participated in our studies and DGs for long years. By using AF method, young doctors gain in knowledge and experience. But high-level clinicians gain also: "after collaboration with high-level mathematicians, the clinician attains a new level of thinking" [12, p7]. This is valuable for teaching purposes. It allows knowledge and experience to pass to the younger generations of doctors.

Acknowledgements

Authors would like to thank Alexandre Vaxman (University of Connecticut, USA) for discussions and help in editing and Jean Roberts (Lancashire School of Health and Post-Graduate Medicine, Preston, UK) for stylistic correction.

References

- [1] M.R.Tonelli. The limits of evidence-based medicine". *Respiratory Care* , 46(12), 2001.
- [2] I.Gelfand, B.Rosenfeld, M.Shifrin, *Essays on collaboration of mathematicians and physicians* (2nd ed.), Moscow, URSS, 2004 (in Russian).
- [3] M.A.Shifrin, O.B.Belousova, E.I. Kasparova. Diagnostic games, a Tool for Clinical Experience Formalization in Interactive "Physician - IT-specialist" Framework". *Proceedings of the Twentieth IEEE International Symposium on Computer-Based Medical Systems*, 2007, 15-20.
- [4] B.M.Reilly, A.T. Evans ,*Translating Clinical Research into Clinical Practice: Impact of Using Prediction Rules To Make Decisions* . *Annals of Internal Medicine*, 2006;144:201-209.
- [5] M.J.Druzdzel, F.J.Diez. Combining knowledge from different sources in causal probabilistic models. *The Journal of machine Learning Research*, vol.4, 2003.
- [6] S.Visweswaran and G.Cooper. Patient-Specific Models for Predicting the outcomes of Patients with Community Acquired Pneumonia. *AMIA Annual Symp Proc*. 2005, 759-763.
- [7] A.Onisko, M.J.Druzdzel, H.Wasylyuk. Learning Bayesian network parameters from small data sets. Application of Noisy-OR gates. *International Journal of Approximate Reasoning*, 27(2), 2001
- [8] M.J.Druzdzel, LC van der Gaag. Buiding probabilistic networks:"Where do the numbers come from?" *IEEE Transaction on Knowledge and Data Engineering*, 12(4), 2000.
- [9] F.J.Diez, J.Mira, E.Iturralde, S.Zubillaga. DIAVAL, a Bayesian expert system for echocardiography. *Artificial Intelligence in Medicine*, 10:59-73, 1997.
- [10] A.Sboner, C.Aliferis Modeling Clinical Judgment and Implicit Guideline Compliance in the Diagnosis of Melanomas Using Machine Learning, *AMIA 2005 Symposium Proceedings* (pp 664-668).
- [11] D.A.Schön, *The Reflective Practitioner* , Basic Books Inc., Printed in USA, 1983.
- [12] I.M.Gelfand, M.L.Izvekova, E.I.Kasparova, A.I.Syrkin. Structural Units in the problem of the prognosis of the Outcome of a Transmural Miocardial Infarction. Preprint, RAN, Moscow 1993 [in Russian].

Disease outbreak detection through clique covering on a weighted ICPC-coded graph

Klaske VAN VUURDEN¹, Gunnar HARTVIGSEN, Johan Gustav BELLIKA
Department of Computer Science, University of Tromsø, Norway

Abstract. Even after a decade of increased research into the problem of detecting disease outbreaks, we lack a system that can limit the number of patients affected by a potential epidemic by recognising its existence at an early stage. In this paper we suggest the use of a weighted graph representing symptoms with an exceptionally high prevalence. Cliques with high weighted edges in such a graph will represent groups of symptoms that occur together more often than usual. As a result each clique will represent the main symptoms of a disease with a high incidence rate. This will make it easier to diagnose the nature of an outbreak, to reach the affected patients at an early stage and to distinguish between outbreaks occurring simultaneously.

Keywords. Epidemiological research, population surveillance, decision support, algorithms, syndromic surveillance

Introduction

In the last decade, many population surveillance systems have been developed with the intention of detecting disease outbreaks [1,2,3]. These systems have in common that they analyse data from a number of people and use statistical methods to make a prediction about the health of a larger population. This paper argues that in order to make a precise prediction about the nature of an outbreak, the accumulated clinical symptoms of a population should be analysed as if they were experienced by a single person. Furthermore, we describe how this principle can be implemented using a weighted undirected graph. If there is sufficient information available about the symptoms of each single patient, it becomes possible to see how different symptoms relate to each other, which in turn can be used to diagnose the population. An identification of an outbreak should then be directly translated into actions towards the current patients and the potential victims of the disease.

Currently, Tromsø Telemedicine Laboratory (TTL) is working on the development of the Snow-system [4,5]. This system will process epidemiological Electronic Health Record (EHR)-data from general practitioners within a region, in order to share knowledge about prevalent contagious diseases. The Snow-system extracts, among others, data in the form of International Classification of Primary Care (ICPC)-codes

¹ Corresponding author: Klaske van Vuurden, Department of Computer Science, University of Tromsø, 9037 Tromsø, Norway. E-mail address: klaske@cs.uit.no

[6,7]. These codes define a classification of symptoms and diagnoses. To diagnose the population of a region as a whole, the sum of all ICPC-data can be searched for groups of symptoms that occur more frequently. Such sets of symptoms could predict what the

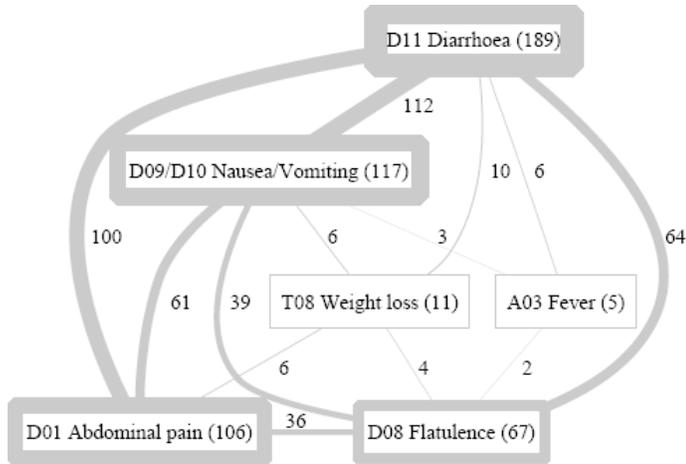


Figure 1. Giardiasis in the city of Bergen, Norway 2004

symptoms of a prevalent disease are. By considering the patient data of an entire region as opposed to that of one individual, there will be less distraction caused by noise. Noise in this context could occur if only a subset of a patient's symptoms is being reported, if symptoms are caused by multiple conditions or if a patient receives a wrong diagnosis. By only looking at groups of symptoms with an exceptionally high frequency, a distinction can be made between different prevalent illnesses in a community.

The detection of groups will be achieved by modelling the distribution of symptoms as a weighted undirected graph, where the nodes represent symptoms, and the edges represent pair wise appearance of symptoms. A *clique* (a set of nodes of which each pair shares a direct relation) with a relatively high weight in such a graph, represents symptoms which are not only prevalent by themselves, but also as a group. When comparing such graphs for different periods in time, incidence rates of different diseases or phases within one disease can be explored. In practice, when cliques with an exceptionally high weight are detected, the community does not necessarily have to be alerted. Any health care professional entering new sets of symptoms into an EHR, can get a report or visualisation informing about the status of that particular set of symptoms. For example, if the set that is entered is a subset of a currently prevalent larger clique, the health care professional could use the other symptoms in that clique as a guideline for further interviewing the patient.

The idea of this system can be sketched by using the example of the outbreak of Giardiasis in Bergen 2004 [8]. The *Giardia Lamblia*-parasite responsible for this disease generally causes symptoms of diarrhoea, nausea, vomiting, abdominal pains, increased flatulence and weight loss. Although the last symptom is less common, a weighted graph of all symptoms would at least reveal higher weights on edges between the nodes representing the first five symptoms. The first suggestion found for these symptoms at Symptom Checker by WebMD [9] is indeed Giardiasis. The situation in

Bergen is illustrated in Figure 1 as based on the symptoms experienced by 200 Giardiasis-patients interviewed by Steen and Damsgaard [8]. The weights on the edges in this figure are the expected number of patients experiencing a pair of symptoms, based on the probability of each individual symptom occurring. The figure displays the situation if the only abnormality occurring was the Giardiasis outbreak. Since the outbreak started in October, it is likely that there was also an increase in the number of influenza patients. The typical symptoms of influenza do not overlap with those for Giardiasis. If we would add data for influenza patients, we could see a distinction between the clique for Giardiasis and that for influenza.

To demonstrate this method on a real situation, we conducted a test on data from the Central Data from General Practitioners (SEDA)-project [10], for which patient records were collected from 20 GP-clinics in Norway during one month in 2001.

The algorithm described here is based on an NP-hard problem, yet an estimation of the maximal cliques should be sufficient to point out the most significant developments in symptoms in a region. Since only the symptoms are analysed and communicated, patients described in the graph can stay relatively anonymous. Therefore, the information extracted from the graph can be used in many different applications.

1. Methods

The symptoms (in the ICPD-classification), as stated in an entry to an EHR, are encoded and sent to a regional server. Using the Snow-system [4], this data will be encoded with a minimal risk to the privacy of patients. At the regional server, a graph $G = (V, E, W_E)$ is maintained representing current sets of symptoms throughout the region. A node v in the graph will represent symptom v and will have a label representing the number of patients with that symptom. An edge (v_1, v_2) will represent the relation between symptoms v_1 and v_2 . The weight $w_{v_1, v_2} \in W_E$ represents the number of patients that have recently reported at least both symptoms v_1 and v_2 . It is also possible, and probably desirable to define and maintain W_V for graph G , representing the number of patients reporting each separate symptom. Adding this information to the maintained data, will not increase the space complexity and more importantly, it could offer interesting options for research. On the other hand, if this would be the only function calculated, the algorithm would fail to identify prevalent groups of symptoms in several cases. If there is a general rise in number of symptoms, it would be difficult to see how they are related and it becomes more difficult to make a diagnosis.

Graph G is a representation of all sets of symptoms that are interesting because they are prevalent. The cliques in graph G will give an approximation of the subsets that are common. Maximal cliques in G can correspond to disease descriptions. In a graph G' representing symptoms with an exceptionally high occurrence (measured against graphs constructed for similar periods), a maximal clique that covers a large amount of patients, can indicate that there is an outbreak of some disease. Indeed, if symptoms A and B both occur more frequently than usual, it is slightly more likely that they occur simultaneously in any individual. If these symptoms are typical for a certain disease, i.e., the symptoms are positively correlated, they will occur together more often when this disease is prevalent.

Ideally, all sets of symptoms occurring should be stored and examined. Collecting this information would lead to the creation of a hypergraph. Such a graph would give

an exact picture of how groups of symptoms occur in a society. However, the storage and analysis of these hypergraphs and comparison between different sets is quite complex. Comparison is necessary because not all patients will display the exact set of symptoms typical for their main condition. The cliques in graph G will give an approximation of the subsets that are common, with a smaller space complexity.

Finding an exact maximum clique is an NP-hard problem [11]. Since we are only looking for approximate sets of symptoms, an algorithm that finds maximal cliques with a high likelihood will be sufficient for the problem discussed in this paper. In the experiment described in the next section, we have used an algorithm that finds some local cliques in linear time. This algorithm makes an approximate covering of the graph by removing detected maximal cliques from the graph until the weights of the graph have reached some minimum. It will be possible that certain false or negative positives occur. To account for this possibility, the results of the algorithm should be handled carefully. The graph should be analysed often, e.g., every minute, and in a random order to look at multiple possibilities. Even if an indication for an outbreak is found, there should be no official alarm until it is confirmed by a health care professional. The weight of a clique should be taken into account as an estimation of the severity of an outbreak. It is theoretically possible that cliques are incorrectly identified as more prevalent. These cases would be rare in real data, and the situations can be overcome by some simple adjustments to the algorithm. Due to the scope of this paper, it will not be possible to prove the correctness of this statement now.

To utilise the algorithm discussed here, we need to establish when a set of symptoms is *more prevalent than usual*. In the next section, we assume that this situation occurs whenever the value of an edge is higher than the measured average value for that same edge. When examining a larger set of data it will become possible to measure fluctuations more accurately and an edge should only be found more prevalent if it lies outside a range of expected values.

2. The SEDA-project

To demonstrate the algorithm as described in the previous section, we implemented a basic version. The algorithm will later on be implemented as part of the Snow-system. At the time this paper was written, the amount of ICPC-data gathered for the Snow-project was not sufficient to create a simulation. Instead, we performed a test on data from the Central Data from General Practitioners (SEDA)-project [10]. This project was conducted in 2001 by Statistics Norway to research the possibility of shared statistics among general practitioners. The data collected for this research includes entries to patient records from 20 GP-clinics in Norway, made in November of that same year. (This is real data and a substitute for the data collected with Snow.)

In most of these records, ICPC-coding was limited to describing the diagnosis. Out of the 30,565 records available, only 23,501 made use of the ICPC-codes and on average only 1.13 codes were used per record. To run the algorithm with the SEDA-data as it would be conducted for ICPC-symptoms, the categories of ICPC (A to Z) were used, instead of all individual symptoms.

We used the first half of the data to obtain information about the normal frequency of relations. Next, we compared each day in the last part of the data to the same weekday in earlier weeks (with each data for one day representing data collected in the last week). Cliques with an increased prevalence of at least 10 cases are shown in Table

1. The results in the table present a prediction about the increased amount of patients affected by conditions with symptoms within certain categories.

Symptoms	Moment of occurrence	Increased number of patients affected
P(Psychological), R(Respiratory)	October 10 – 18, 2001	at least 10
R(Respiratory), S(Skin)	October 10 – 18, 2001	at least 10
T (Endocrine, metabolic and nutritional), L (Musculoskeletal), K (Cardiovascular)	October 18 – 28, 2001	at least 15
A(General), X(Female Genital)	October 20 – 27, 2001	at least 10

Table 1. Sets of symptom/diagnosis-categories with increased weight in the SEDA-data

3. Discussion

The results in Table 1 show that the algorithm described in this paper can detect cliques of prevalent symptoms from the accumulated symptoms of all symptoms reported by patients over a period of time. We cannot diagnose these results yet, since the symptom-groups are too general. However, when inspecting for example the clique of A and X-symptoms more carefully, we can conclude that the amount of pregnant women with general complaints was high in the period from the 20th to the 27th of October.

We need a data set with a higher average number of ICPC-codes to find more interesting results. To fully utilise the algorithm described in this paper, all symptoms described by patients should be documented. A possibility to overcome this problem in the Snow-system is described by Johansen et al. [4]. They propose to let patients enter their symptoms in a system while still in the waiting room. This will give patients the opportunity to describe their symptoms thoroughly without causing more workload for the physician. This method will likely cause an increase in the average number of symptoms per patient record.

It also needs to be noted here that one month of data is very little to extract reliable knowledge about patterns in the occurrence of different symptoms. Since the Snow-project will run for an indefinite period, it will likely provide an increased number of ICPC-symptoms. Experiments conducted on the data set collected by the Snow-system will verify whether the statements made in this paper hold for real data with real noise. The Snow-system will also provide us with an overview of actual occurrences of disease outbreak, which will make it possible to verify our findings.

Once implemented, the proposed method is non-intrusive. Only a minimal amount of data is needed (if only symptoms are used, the identity of individual patients is very difficult to trace) and data is only used to be inserted in the graph. The graph itself is not stored for each moment in time; it is only integrated in data representing patterns in the graph over different time periods. As a result, data will become more and more abstract over time. To get a complete picture of the health status of a community, the graph could be extended to describe diagnoses made or demographic information about patients. This would, however, make the data less anonymous and some relations detected could represent sensitive information.

The methods described in this paper could potentially be used for many different purposes. In the introduction, we touched upon the possibility of decision support when

adding a new set of symptoms to the regional graph. This implementation will provide physicians with statistical data about the current situation in their region, which can assist them in diagnosing. Other purposes could be automatic detection, visualisation of symptoms and diseases within a region, distinction between rare and common symptoms of a disease or identification of early symptoms caused by a disease. Another purpose could be that of integrating graphs for different regions with each other to form information about the global spread of pandemics. All of these possibilities could be further explored once a simple version is implemented as part of the Snow-system.

4. Conclusions

This paper examined the possibility of exploring a weighted graph in which the nodes represent clinical symptoms or diagnoses and where the edges represent pair-wise relations between the symptoms occurring in a certain region. Though the algorithm proposed is based on a complex problem, we are only searching after an estimation of all maximal cliques. With an algorithm that can approximate maximal cliques, there is a high probability that prevalent cliques will be found. With these cliques a prediction can be made about the nature of diseases currently prevalent and a distinction can be made between different independent diseases. This method will not only provide us with information about the occurrence of a disease outbreak, but it will also give an outline of the set of symptoms associated with the disease. The population affected by the disease can be approached directly and the process of diagnosing the actual disease can be accelerated.

5. Acknowledgments

This paper was partly funded by the Tromsø Telemedicine Laboratory (TTL) project 174934. We would like to thank all researchers at TTL involved in the development of the Snow-system.

References

- [1] D. L. Buckeridge. Outbreak detection through automated surveillance: A review of the determinants of detection. *Journal of Biomedical Informatics*, 40:370–379, 2007.
- [2] D. M. Bravata, K. M. McDonald, W. M. Smith, C. Rydzak, H. Szeto, D. L. Buckeridge, C. Haberland, and D. K. Owens. Systematic Review: Surveillance Systems for Early Detection of Bioterrorism-Related Diseases. *Annual Internal Medicine*, 140:910–922, 2004.
- [3] M. M. Wagner, F.-C. Tsui, J. U. Espino, V. M. Dato, D. F. Sittig, R. A. Caruana, L. F. McGinnis, D. W. Deerfield, M. J. Druzdel, D. B. Fridsma. The emerging science of very early detection of disease outbreaks. *Journal of Public Health Management and Practice*, 7:51–59, 2001
- [4] J. G. Bellika, T. Hasvold, and G. Hartvigsen. Propagation of program control: A tool for distributed disease surveillance. *International Journal of Medical Informatics*, 76(4):313–329, 2007.
- [5] M. A. Johansen, G. Aronsen, G. Hartvigsen, and J. G. Bellika. What do GPs want in return from a syndromic surveillance system? *Advances in Disease Surveillance*, 4:168, 2007.
- [6] International Classification Committee of WONCA. *ICPC-2 International Classification of Primary care*. Oxford University Press, 2 edition, 1998.
- [7] I.M. Okkes, M. Jamouille, H. Lamberts, and N. Bentzen. ICPC-2-E: The electronic version of ICPC-2. Differences with the printed version and the consequences. *17(2):101–6*, 2000.
- [8] K. Steen and E. Damsgaard. Giardiaepidemien i 2004 og Bergen Legevakt. *Tidsskrift for Den Norske Lægeforening*, 127(2):187–189, 2007.
- [9] Web MD Inc. Symptoms checker. <http://symptoms.webmd.com/symptomchecker>, October 28, 2007.
- [10] A. Sundvoll and I. Kvalstad. SEDA - Sentrale data fra allmennehetjenesten, Sluttrapport fra pilotprosjektet. http://www.ssb.no/emner/03/02/20/rapp_200213/, 2002.
- [11] R. M. Karp. Reducibility Among Combinatorial Problems. *Complexity of Computer Computations*,

4. Evaluation

This page intentionally left blank

A Randomized Controlled Trial to Evaluate an Electronic Scoring Tool in the ICU

Thomas BÜRKLE^{a1}, Alexander BEISIG^a, Marion GANSLMAYER^b,
Hans-Ulrich PROKOSCH^a

^a*Chair Medical Informatics, University Erlangen, Germany*

^b*Department Internal Medicine 1, University Hospital Erlangen, Germany*

Abstract. Few RCTs on the effect of computer applications in intensive care have been published. This study presents an RCT measuring time savings and score values after introduction of a computer based scoring tool in an intensive care unit. A tablet PC with a standalone scoring application for TISS, SAPS2 and Apache 2 was supplied. We measured considerable time savings and higher score values when the computer application is used.

Keywords. RCT, Information system, Intensive Care

Introduction

When the influential 1999 IOM report “To Err is human: Building a Safer Health System” [1] was published, it prompted considerable discussion how better patient care could be achieved by using more computer applications. There have been critical reports however, that before stating that we could save so many lives we would need to thoroughly evaluate our computer applications and to prove that they fulfill the expectations [2]. However, it is still difficult to implement good evaluation studies for computer applications [3] and randomized controlled studies (RCTs) evaluating computerized information systems remain rare [4]. There have been many publications about electronic support and patient data management systems (PDMS) in intensive care units ICU [e.g. 5,6] but only few RCT’s on this topic [7,8,9,10]. A Medline search refined for RCTs with the MeSH terms “Information Systems” and “Intensive Care” revealed only six publications.

This paper presents an RCT at the medical intensive care unit 1 (MICU-1) of Erlangen University Hospital, which was performed to evaluate a standalone scoring application on a tablet PC. Our goals was to assess user satisfaction (by means of a questionnaire), to find out how workflows changed (by workflow analysis), to measure time needed to score a patient (by RCT) and to compare score values generated with help of computer and without (by RCT).

¹ Corresponding Author: PD. Dr. Thomas Bürkle, Lehrstuhl für Med. Informatik, Krankenhausstrasse 12, D 91054 Erlangen, Germany, Email: Thomas.Buerkle@imi.med.uni-erlangen.de

1. Environment

Our university hospital has several ICUs. MICU-1 belongs to the department Internal Medicine 1 which specializes for gastroenterology, pneumology and endocrinology. MICU-1 has 12 beds, 44 nurses and 11 physicians in rotational shifts. In 2005 they cared for 537 patients and delivered more than 48.000 respirator hours and 499 days of CVVH. Therapy comprises diseases like ARDS, pneumonia, sepsis, liver failure, gastrointestinal bleeding, severe clotting disorder, poisoning, metabolic failure as well as post-transplantation complications, usually combined with multi organ failure. In contrary to some other Erlangen University Hospital ICUs there is currently no PDMS available on MICU-1. At the time of this study the computing infrastructure of MICU-1 comprised four workstations in the physician office and two more in the nursing office to access e.g. the Erlangen Hospital Information System HIS (Siemens Soarian) and to perform lab order entry (Swisslab) and radiology order entry (iSoft RadCentre). The workstations in the nursing office support electronic drug ordering from the pharmacy. There is a central monitoring system for all beds at the nursing office (Siemens Infinity). Furthermore a PACS workstation and printer are available for retrieval of radiology images (Siemens). The patient chart and patient record however are maintained completely on paper.

German legal and reimbursement requirements [11] necessitate since 2006, that for each intensive care patient the scores CORE10TISS, a TISS28 derivate (therapeutic intervention scoring system [12,13]) and SAPS2 (simplified acute physiology score [14]) are collected daily. Hospitals must supply accumulated score sums for each ICU patient. The totaled score is mapped to the German OPS-Code (an ICPM derivate) 8-980 which stands for complex intensive care treatment. In combination with the totaled respirator hours this data is used for reimbursement. Traditionally and for scientific evaluations MICU-1 collects daily two other scores for each patient, namely the APACHE2 [15] and the TISS76 [16] score. Before this study, all scoring was done manually on paper and the results entered in the paper based patient chart and a small MS-Access database application for statistical evaluation. This was a time consuming task.

2. Methods

On request, the medical informatics department analyzed the current situation and documentation workflow at MICU-1 and offered to implement a computerized scoring tool as an interim solution as long as no PDMS is available. The application was designed for tablet PC to support bedside score documentation and programmed in JAVA using Hibernate 3.2 and HSQL 1.8.0 as database and the Apache FOP library for graphical output. It's specifications comprised download compatibility into the existing Access database, easy configuration of new scores with XML template, and user profiling to permit each user to define his own sequence of data entry. Application development was based on rapid prototyping in close collaboration with clinicians. The scoring tool went into use in December 2006. Figure 1 shows the user interface which is optimized for pen entry.

To evaluate the new scoring tool a complex evaluation protocol was defined combining workflow analysis, time series questionnaire technology and a randomized trial measuring time consumption and score values. Here we will concentrate on the

RCT results to answer the hypotheses “the system has changed the time needed to document scores” and “the system has changed the quality of score documentation”. To support or reject these hypotheses n=54 patient scoring events performed in March and April 2007 have been groupwise randomized into 27 interventions where the computer application was used for scoring and 27 controls where scoring was performed manually as before. A scoring event is defined as collecting four scores CORE10TISS, SAPS2, APACHE2 and TISS76 once at midnight for one patient. Time measurement was done manually using a stop watch. Measurement started when the paper based patient chart (plus tablet PC in intervention) were available and finished when the respective score values were written in the patient chart. In the manual group each of the four scores was timed separately, whereas in the intervention group we measured time for SAPS2 and APACHE2 combined, because the documentation workflow had been streamlined in the computer application. There, physiology parameters for both SAPS2 and APACHE2 could be documented just once and the program did allocate the respective score values for both (see Figure 1). For total time measurement, times to fetch or distribute the paper patient records were included.

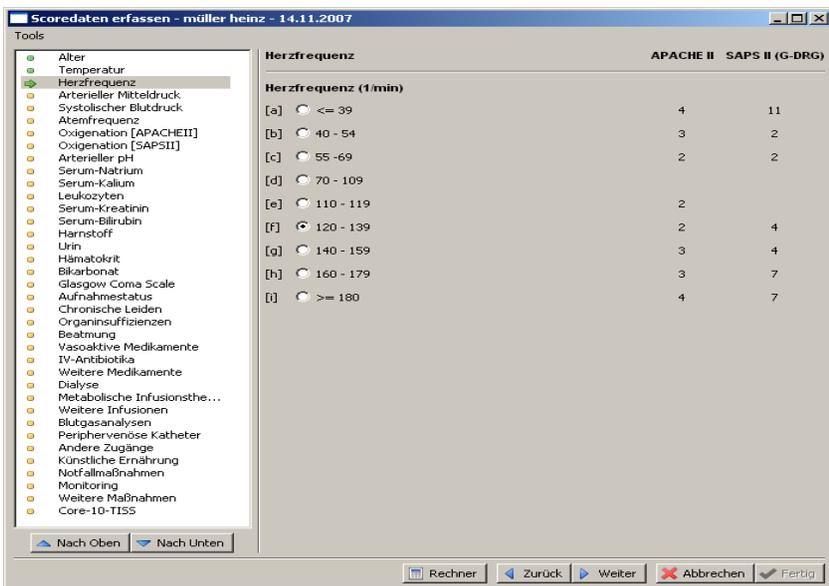


Figure 1: data entry screen of score tool on tablet PC

Furthermore we collected all score values from intervention and control group. In order to assess the quality of the documented score values, all scores of both groups have been collected a second time from the researcher (author AB) using the computer tool. Then measurements have been collected a third time from a senior physician of the ward, thus establishing the presumably correct gold standard.

For statistic testing we used Mann Whitney U test assuming 5% error level.

3. Results

Median values of scoring time measurements in seconds are shown in Figure 2. Documentation time for the combination of SAPS2 and APACHE2 (both times have been summed in the manual group) is nearly halved from 150 to 80 seconds with computer assistance. For TISS76 the time saving is 60 seconds on computer compared to 95 seconds on paper, for CORE10TISS it is 10 against 15 seconds and the total time is 165 seconds against 265 seconds. All time differences are significant at the 5% level.

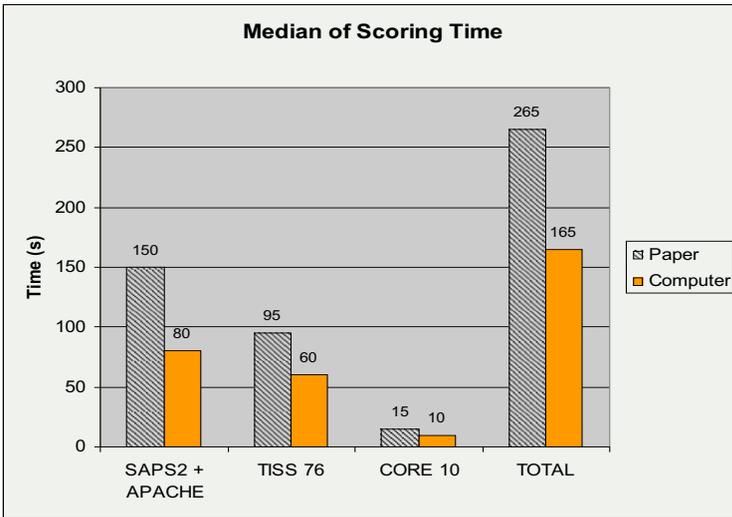


Figure 2. Median time values for scoring using paper or computer

Figure 3 shows score results for APACHE paper versus gold standard on the left (27 randomized score values measured twice) and scoring tool versus gold standard (another 27 randomized score values measured twice).

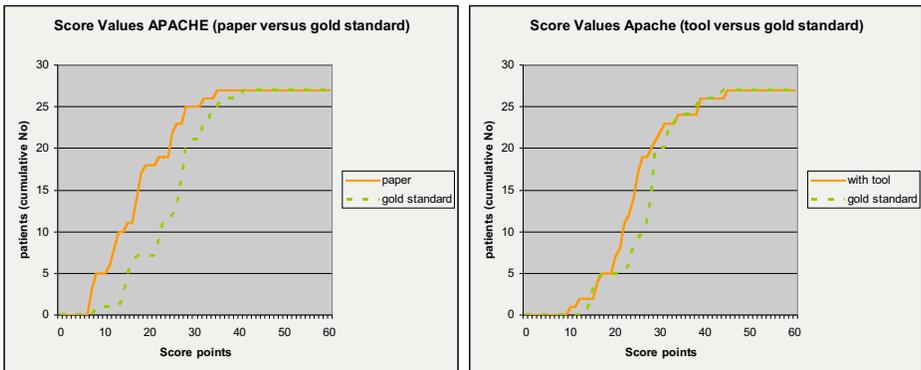


Figure 3. Score values compared to gold standard, control (paper based scoring) left, intervention (computer tool) right

APACHE2 scores on paper (Figure 3 left) were 10 points lower than gold standard (significant at 5%), compared to scores in the intervention group with computer (Figure 3 right) which were only 4 points lower than gold standard (non significant, $p=6.4\%$). Accordingly we found a difference of 6 points to the gold standard for SAPS2 when using paper (significant) compared to only 4 points difference when using computer (non significant). For TISS76, values using paper were 4 points lower than gold standard (significant) while computer score values were 2 points higher than gold standard (non significant). For CORE10TISS there were no significant differences between measurement and gold standard neither in intervention nor in control group. In summary we note a significant decrease of scoring time when using computer and higher score values near to our gold standard for scoring APACHE2, SAPS2 and TISS76 on computer.

4. Discussion

Within this evaluation we could demonstrate a significant decrease in time used for scoring intensive care patients by introducing a plain computer based scoring tool. These results correspond nicely with the results of our user satisfaction evaluation which is beyond the scope of this paper. We identified one other RCT [7] which did also demonstrate time savings for computerized documentation, but did so for a full grown commercial PDMS and measured nursing activities. We measured physicians work time and our intervention was only a minor stand alone application taking less than 3 man months for development.

Furthermore we could demonstrate a significant influence of computerized scoring on scoring quality, since manual score values for three out of four scores were too low. It remains difficult to establish a valid gold standard for scoring ICU patients. One physician may e.g. consider a documented high pulse or pressure value from a monitor as an artefact due to patient movements whereas another colleague may trust this value because he had attended the measurement himself and therefore is confident that the peak reflects reality. Therefore our score values are not fully identical between gold standard (established by repeated measurements) and both computer scoring or paper scoring. It remains a fact however, that those differences are small and insignificant when computerized scoring was used, whereas values were significantly lower when paper was used, except for the elementary CORE10TISS score. There is one other study of Bosman and colleagues [17] who performed simultaneous manual and computerized charting using a PDMS and note, similar to our study, higher severity scores for APACHE2, SAPS2 and MPM (Mortality Probability Models) resulting in an increased predicted mortality. They do however not use a gold standard such as repeated scoring to confirm which values are better. In that study, physiology parameters were drawn directly from the PDMS for computerized scoring and from the paper record for manual scoring, whereas our source information was always a paper based patient record. Therefore we conclude that even without direct and automated transmission of vital signs into an electronic patient chart a computerized documentation tool may alter scoring results.

However, critique to this evaluation study should be allowed. This is a small study with $n=54$ scoring events. The designer of the application took part in the evaluation of his own program, which could lead to a Hawthorne effect. In a disruptive environment such as an intensive care unit, there are many interfering factors which may influence

an interventional study. We tried to overcome both facts by hopefully distributing interfering factors equally among intervention and control group using a rigid study protocol and by measuring differences against a gold standard.

Our results suggest that ICU scoring of non trivial scores should always be done with computer support, not only to save time, but to achieve a higher quality of score values. Considering the scoring application itself we are fully aware that we have dealt only with one minor documentation aspect on MICU-1. We plan to support this ward in future with a commercial PDMS which will draw vital signs for scoring SAPS2 directly from imported patient monitor data. But our results indicate, that we should increase computerized documentation in ICUs for better data quality, despite of the risks which arise from increased dependency of computer applications.

References

- [1] L. Kohn, J. Corrigan, M. Donaldson (eds), *To Err is Human: Building a Safer Health System*. Washington DC: National Academic Press, 1999.
- [2] Institute for Safe Medication Practices, *Over-reliance on computer systems may place patients at great risk. ISMP Medication Safety Alert* Huntingdon Valley, Pa.: ISMP, 1999.
- [3] T. Bürkle, E. Ammenwerth, H.U. Prokosch, J. Dudeck, Evaluation of clinical Information Systems - What can be evaluated and what cannot? *J Eval Clin Pract.* **7(4)** (2001), 373-85.
- [4] E. Balas, S. Austin, J. Mitchel, B. Ewigmann, K. Bopp, G. Brown, The clinical value of computerized information services. A review of 98 randomized clinical trials. *Archives of Family Medicine* **5** (1996), 271-278.
- [5] T. Bürkle, A. Michel, W. Horch, L. Schleifenbaum, J. Dudeck, Computer based Nursing Documentation - Means to achieve the goal. *Int J Med Inf.* **52(1-3)** (1998), 71-80.
- [6] R. Röhrig, S. Böttger, A. Junger, M. Benson, A. Michel, G. Hempelmann, Vollautomatische Score-Evaluierung mit einem Patienten-Daten-Management-System (PDMS). *Anästhesiologie & Intensivmedizin* **42** (2001), 409-477.
- [7] R. Bosman, E. Rood, H. Oudemans-van Straten, J. Van der Spoel, J. Wester, D. Zandstra, Intensive care information system reduces documentation time of the nurses after cardiothoracic surgery. *Intensive Care Med* **29** (2003), 83-90.
- [8] E. Rood, R. Bosman, J. van der Spoel, P. Taylor, D. Zandstra, Use of a computerized guideline for glucose regulation in the intensive care unit improved both guideline adherence and glucose regulation. *JAMIA* **12,2** (2005), 172-180.
- [9] S. Weingarten, M. Redinger, L. Conner, T. Lee, I. Hoffman, B. Johnson, G. Ellrodt, Practice guidelines and reminders to reduce duration of hospital stay for patients with chest pain: an interventional trial. *Annals of Internal Medicine* **120(4)** (1994), 257-263.
- [10] K. Colpaert, B. Claus, A. Somers, K. Vandewoude, H. Robays, J. Decruyenaere, Impact of computerized physician order entry on medication prescription errors in the intensive care unit: a controlled cross sectional trial. *Critical Care* **10(1)** (2006), R21
- [11] B. Rapp, DRG-Update Intensivmedizin 2006. *Das Krankenhaus* **1** (2006), 32-39.
- [12] D. Cullen, M. Civetta, B. Briggs, L. Ferrara, Therapeutic intervention scoring system: a method for quantitative comparison of patient care. *Crit Care Med* **2** (1974), 57-60.
- [13] D. Miranda, A. de Rijk, W. Schaufeli, Simplified Therapeutic Intervention Scoring System: the TISS-28 items--results from a multicenter study. *Crit Care Med.* **24(1)** (1996), 64-73.
- [14] J. Le Gall, S. Lemeshow, F. Saulnier, A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* **270(24)** (1993) 2957-2963.
- [15] W. Knaus, E. Draper, D. Wagner, J. Zimmerman, APACHE II: A severity of disease classification system. *Crit Care Med* **13** (1985), 818-829.
- [16] R. Keene, D. Cullen, Therapeutic Intervention Scoring System Update 1983. *Crit Care Med* **11** (1983), 1-3.
- [17] R. Bosman, H. Oudemans-van Straten, D.F. Zandstra. The use of Intensive care information systems alters outcome prediction. *Intensive Care Med* **24** (1998), 953-58.

Evaluating the impact of a service-oriented framework for healthcare interoperability

Stylios DASKALAKIS^{a1} and John MANTAS^b
^{ab}*Health Informatics Laboratory, Faculty of Nursing,
National and Kapodistrian University of Athens, Greece*

Abstract. This paper describes the evaluation of a service-oriented prototype implementation. The prototype development aims to exploit the use of service-oriented concepts for achieving healthcare interoperability while it also attempts to move towards a virtual patient record paradigm. The proposed evaluation strategy investigates the adaptation of the DeLone and McLean model of information systems success with respect to service-oriented implementations. Specific service-oriented and virtual patient record characteristics were empirically encapsulated in the DeLone and McLean model and respective evaluation measures were produced. The proposed theoretical framework was utilized for conducting an empirical study amongst sixty two participants in order to observe their perceptions with respect to the hypothetical adoption of the prototype framework. The data gathered was analyzed using partial least squares. The generated results highlighted the importance of information quality whereas system quality did not prove to be a strong significant predictor in the overall model.

Keywords. interoperability, process, assessment-evaluation: IS success model, DeLone and McLean, partial least squares

Introduction

The evolution of technology in the healthcare context consequently leads to both experimental and productive adoptions of new information technology (IT) standards in the healthcare domain. A typical field of health informatics related with such adoptions is interoperability. Healthcare interoperability includes several aspects under consideration since it is associated with a variety of concerns, perceptions and approaches whereas it also involves several standards, methods, stakeholders and roles [1].

A service-oriented prototype implementation was developed at the health informatics laboratory, Faculty of Nursing, University of Athens [1]. The aim of the prototype implementation is twofold. Primarily to investigate the adaptation of service oriented architectures (SOA) [2-4] for achieving healthcare interoperability. Secondly, to propose a homogeneous Web-based environment, capable of presenting the dynamic

¹ Corresponding author: Stylios Daskalakis, Health Informatics Laboratory, Faculty of Nursing, National and Kapodistrian University of Athens, 123 Papadiamantopoulou street, Goudi 115 27, Athens, Greece; Email: stdask@nurs.uoa.gr.

unification of distributed patient data, through the orchestration of services from a diversity of applications. Consequently, the prototype implementation attempts to formulate a virtual patient record (VPR) approach [1]. Moreover, the research work exploits the use of the business process execution language (BPEL) standard for the design and implementation of business and technical processes that coordinate software services in order to achieve interoperability in healthcare organisations [2,5].

Related work in this field includes initiatives and propositions for VPR implementations [6,7] and service-oriented computing in healthcare [3,4,8,9]. However, equal attention should be paid in the evaluation of technological adoptions in healthcare, in order to ensure the success and effective use of any information system development [10]. The work described in this paper aims to move towards an evaluation strategy for SOA implementations with VPR functionality in the healthcare domain.

The presented work is organised in several sections. Section 1 outlines the research model and method. Section 2 focuses on the data analysis approach. Section 3 presents the results obtained and provides a discussion on the methodology in terms of the measurement and the structural model. Section 4 provides a discussion on the results and the outcomes of the study. Finally, Section 5 presents future adoptions of the SOA prototype framework and the proposed evaluation approach.

1. Research model and method

1.1. Theoretical framework

The theoretical background for conducting an evaluation of the SOA prototype is based on the initial proposition of the DeLone and McLean information systems (IS) Success model [11]. The current study attempts to utilize the IS Success model in the context of a hypothetical adoption, since the SOA framework under evaluation is a prototype. Regardless of the empirical encapsulation of SOA and VPR characteristics within the dimensions of the DeLone and McLean model, the objective is to observe and comment on the trend and the perception that such a prototype might create amongst participants with clinical and/or informatics background.

In their model, DeLone and McLean define six dimensions namely system quality, information quality, use, user satisfaction, individual impact and organisational impact [11]. By definition, the dimensions of system and information quality both affect the dimensions of use and user satisfaction respectively. Furthermore, use and user satisfaction affect each other while both also affect the dimension of individual impact and finally, individual impact affects organisational impact [11-14]. In the context of the current study, the characteristics that contribute to the six dimensions defined by DeLone and McLean are formulated from a SOA and VPR perspective, combined with various relevant characteristics from the influential studies of Iivari [12] and Yusof, Paul and Stergioulas [15]. Thus, for each dimension, specific characteristics are proposed in an attempt to reflect service-oriented and virtual patient record attributes.

1.2. Service-oriented characteristics and hypotheses

Regarding system quality, the selected characteristics were flexibility, usefulness, interoperability, maintainability, scalability and reusability. The selection of the first

three attributes was influenced from the propositions of Iivari [12] and Yusof et al. [15] whereas the last three are proposed in an attempt to reflect possible SOA characteristics and architectural advantages [1,2,5]. Concerning information quality, characteristics include data completeness, data unification, semantically mapping capabilities, consistency of the actual content, data standardization and relevancy of the consolidated data. In a similar manner, the selection of completeness, consistency and relevancy was influenced from Yusof et al. [15] and Iivari [12]. The remaining attributes were selected in an attempt to reflect possible VPR characteristics. Continuously, the rest of the dimensions, namely use, user satisfaction, individual impact and organisational impact, were formulated based on a subjectivist manner.

The dimension of use is related with attributes that attempt to measure the potential use of the prototype, due to its hypothetical nature. Furthermore, for user satisfaction, the current study attempts to include characteristics related with the understanding and the possible innovative nature of the overall architecture, the interest on the proposed model, and finally the perception regarding its capabilities for achieving interoperability. The dimension of individual impact introduces measures related with possible time saving for accomplishing individual tasks, harmonization of daily activities with the overall business process model in a healthcare organisation, usefulness of a SOA framework as both a business and technical tool and finally, individual perceptions regarding productivity and work performance improvement. At last, the dimension of organisational impact formulates a set of attributes that are associated with possible cost implications, business process (re) engineering and/or (re) design along with possible overall time savings from an organisational perspective, in an attempt to investigate such concepts.

The following hypotheses were formulated: System quality affects user satisfaction (H1), information quality affects user satisfaction (H2), system quality affects use (H3), information quality affects use (H4), use affects user satisfaction (H5), user satisfaction affects use (H6), user satisfaction affects individual impact (H7), use affects individual impact (H8) and individual impact affects organisational impact (H9) [12].

2. Data analysis

The prototype nature of the proposed framework creates several limitations. Thus, the sample selection for conducting an evaluation had a limited scope. Mainly, it was based on past and present postgraduate students specializing in health informatics and healthcare services management at the Faculty of Nursing, University of Athens. The undergraduate background of the students was nursing or informatics science. Also, a minor set of participants originated from either nursing departments of healthcare organisations or the healthcare IT industry, and agreed to participate in the study in an individual basis. The empirical assessment included sixty two participants. A significant percentage of the sample (35.5%) was employed in healthcare organisations.

The formulated sample provides the opportunity to obtain hypothetical, but still useful beliefs and perceptions regarding the proposed framework. A short introduction on design issues and a live demonstration scenario of the prototype were presented to the evaluators. Afterwards, participants were asked to complete an anonymous questionnaire. System quality, information quality, use, user satisfaction, individual impact and organisational impact were related with SYSQ (24 measures), INFQ (24 measures), PU (2 measures), PUS (6 measures), INDV (6 measures) and ORG (6

measures) accordingly. All measures were modeled in a 7-scale Likert approach. In general, the number of measures per dimension and the scaling in the questionnaire were adopted from Iivari [12].

The data gathered was analyzed using structural equation modeling, specifically partial least squares (PLS) which is mainly used for prediction rather than confirmatory analysis [12,17] and is utilized in several research works [12,13,18-20]. For the modelling and the data analysis, the software package smartPLS version 2.0 M3 beta was used [16]. The dimensions and characteristics were described as a second order model with two scenarios in smartPLS software, in order to adequately reflect the bi-directional interaction between the dimensions of use and user satisfaction (use to user satisfaction and user satisfaction to use respectively) [12].

3. Results

The generated results were assessed according to the measurement and the structural model [12, 13, 17-20].

For the measurement model, individual item loadings, construct reliability (internal consistency), convergent validity and discriminant validity were investigated. Regarding individual item loadings, 66 out of 68 overall items exceeded 0.6, with 61 of them producing a value greater than 0.7, thus considered reliable [12,13,18-20]. Only two items produced values below 0.6 in both scenarios, but can be considered acceptable since they exceed the threshold value of 0.5 for acceptable results [19,20]. Construct reliability assessment was based on Cronbach's alpha and composite reliability. All values were considered reliable (exceeded 0.7 for Cronbach's alpha and 0.8 for composite reliability) [12,13,18-20], except the value of Cronbach's alpha for use (0.5976). In the case of convergent validity, the values of the average variance extracted (AVE) were observed, as proposed by Fornell and Larcker [21] [12,13,19]. All constructs produced an AVE that exceeded 0.5 [12,13,18-20], except system and information quality (0.4048 and 0.4298 respectively), which may be explained from the molar nature of the model [12]. At last, for the assessment of discriminant validity, all constructs produced satisfactory results for the square root of the AVE [12,13,18-20], except interoperability (0.7792 compared to 0.8459), scalability (0.8004 compared to 0.8233 and 0.8236), flexibility (0.7924 compared to 0.806 and 0.8063) and reusability (0.7973 compared to 0.8602). Additionally, system and information quality did not produce satisfactory results (0.6362 compared to 0.7678 and 0.7721 for system quality and 0.6555 compared with 0.8504 and 0.8506 for information quality in both scenarios).

Regarding the structural model, a resampling technique was used, based on bootstrapping (500 resamples) [12,13,17,18,20]. The examination of the t-values was based on a two-tail test with statistically significant levels of $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.001$ (***). The results are shown in Figure 1. The upper value per pair concerns the first scenario whereas the lower value the second scenario respectively. Dotted lines emphasize the paths that did not prove to be significant in both scenarios and consequently the hypotheses that were not confirmed.

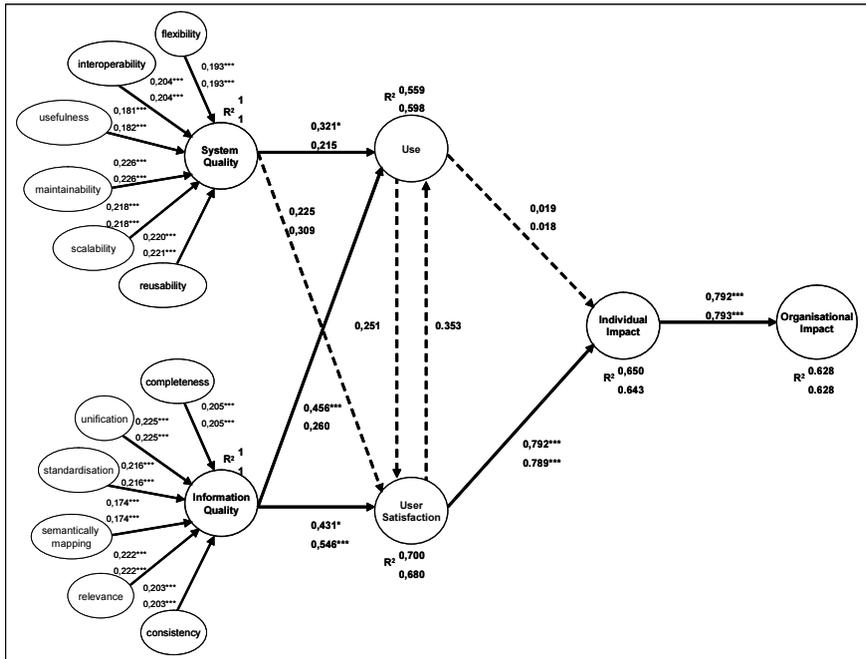


Figure 1. Results of the structural model (Source: Adopted from Daskalakis [22])

4. Discussion

The evaluation results highlighted that system quality is partially a predictor of system use but not of user satisfaction. Moreover, system use did not prove to be a significant predictor of individual impact. On the contrary, information quality proved to be a significant predictor of user satisfaction and partially a significant predictor of system use as well. Furthermore, the bi-directional relation between use and user satisfaction did not confirm in the context of the current study. In addition, user satisfaction found to be a strong predictor of individual impact. Finally, the dimension of individual impact was found to be a strong significant predictor of organisational impact. Such findings may highlight that participants tend to focus more on the quality of the information provided than the characteristics that govern the technology used for the information provisioning.

5. Conclusions

A service-oriented interoperability prototype with VPR capabilities has been developed and evaluated, based on a hypothetical adoption. Future work in the current field includes the prototype adoption in real healthcare conditions. Its impact can be investigated by conducting empirical assessments with distinct stakeholder groups. Overall, the propositions described in the current paper may be potentially adjusted in a variety of healthcare organisations in order to promote the evaluation of service-oriented implementations with virtual patient record capabilities.

References

- [1] S. Daskalakis and J. Mantas, A SOA framework for healthcare interoperability using BPEL: Design and implementation, *Proceedings of the 4th International Conference on Information and Communication Technologies in Healthcare*, Samos, Greece, (2006), 247-254.
- [2] M.P. Papazoglou, Service-oriented computing: Concepts, characteristics and directions, *Proceedings of the 4th International Conference on Web Information Systems Engineering (WISE'03)*, Rome, Italy, (2003), 3-10.
- [3] D.W. Forslund, J.E. George, S. Koenig, T. Staab, M. Kratz and R. Carter, The importance of distributed, component-based healthcare information systems: The role of a service-based architecture, *Proceedings of the 14th IEEE Symposium on Computer-Based Medical Systems*, Bethesda, MD, USA, (2001), 79-82.
- [4] W. Wang, M. Wang and S. Zhu, Healthcare information system integration: A service oriented approach, *Proceedings of the 2005 International Conference on Services Systems and Services Management (ICSSSM'05)*, Chongqing, China, (2005), 1475- 1480.
- [5] Z. Zou and Z. Duan, Building business processes or assembling service components: Reuse services with BPEL4WS and SCA, *Proceedings of the 4th European Conference on Web Services (ECOWS'06)*, Zurich, Switzerland, (2006), 138-147.
- [6] F. Malamateniou and G. Vassilacopoulos, Developing a virtual patient record using XML and web-based workflow technologies, *International Journal of Medical Informatics* **70** (2003), 131-139.
- [7] A. Berler, S. Pavlopoulos, G. Karkalis, E. Sakka, G. Konnis and D. Koutsouris, Implementation of a novel virtual patient record architecture, *Proceedings of the 24th Annual International Conference of the IEEE in Medicine & Biology*, Houston, TX USA, (2002).
- [8] S. Chu, From component-based to service oriented software architecture for healthcare, *Journal on Information Technology in Healthcare* **4** (2006), 5-14.
- [9] R. Lenz, M. Beyer and K.A. Kuhn, Semantic integration in healthcare networks, *International Journal of Medical Informatics* **76** (2007), 201-207.
- [10] E. Ammenwerth, J. Brender, P. Nykänen, H-U. Prokosch, M. Rigby and J. Talmon, Visions and strategies to improve evaluation of health information systems: Reflections and lessons based on the HIS-EVAL workshop in Innsbruck, *International Journal of Medical Informatics* **73** (2004), 479-491.
- [11] W.H. DeLone and E.R. McLean, Information systems success: The quest for the dependent variable, *Information Systems Research* **3** (1992), 60-95.
- [12] J. Iivari, An empirical test of the DeLone-McLean model of information system success, *The data base for Advances in Information Systems* **36** (2005), 8-27.
- [13] J.L. Roldán and A. Leal, A validation test of an adaptation of the DeLone and McLean's model in the Spanish EIS Field, In *Critical reflections on information systems: a systemic approach*. J. J. Cano. IGI Publishing, 2003.
- [14] A. Molla and P.S. Licker, E-commerce systems success: An attempt to extend and respecify the DeLone and McLean model of IS Success, *Journal of Electronic Commerce Research* **2** (2001), 131-141.
- [15] M.M. Yusof, R.J. Paul and L.K. Stergioulas, Towards a framework for health information systems evaluation, *Proceedings of the 39th Annual Hawaii International Conference on System Sciences*, Big Island, Hawaii, USA, (2006), 95a.
- [16] C.M. Ringle, S. Wende and A. Will, SmartPLS 2.0 (M3) beta, Available from: <http://www.smartpls.de>, University of Hamburg, 2005.
- [17] D. Gefen, D.W. Straub and M. Boudreau, Structural equation modeling techniques and regression: Guidelines for research practice, *Communications of AIS* **4** (2000).
- [18] S. Chea and M.M. Luo, Cognition, emotion, satisfaction and post-adoption behaviors of e-service customers, *Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, Big Island, Hawaii, USA, (2007), 154b.
- [19] M.M. Shepherd, D.B. Tesch and J.S.C. Hsu, Environmental traits that support a learning organization: The impact on information system development projects, *Comparative Technology Transfer and Society* **4** (2006), 196-218.
- [20] O. Yigitbasioglu, Information sharing with key suppliers: A contingency approach, *Proceedings of the 10th European Conference on Accounting Information Systems*, Lisbon, Portugal, (2007).
- [21] C. Fornell and D.F. Larcker, Evaluating structural equation models with unobservable variables and measurement error, *Journal of Marketing Research* **18** (1981), 39-50.
- [22] *Application integration and interoperability of healthcare information systems*. S. Daskalakis. PhD Dissertation, National and Kapodistrian University of Athens, 2007.

The use of performance metrics to monitor the impact of CPOE on pathology laboratory services

Andrew GEORGIU^{a,1}, Wendy MORSE^b, Wyndham TIMMINS^b, Sangeeta RAY^c,
Johanna I. WESTBROOK^a

^a *Health Informatics Research & Evaluation Unit, Faculty of Health Sciences, The University of Sydney, 1825, Australia*

^b *Sydney South West Area Health Service Pathology Liverpool, 1871, Australia*

^c *Discipline of International Business, The University of Sydney, 2006 Australia*

Abstract. Organisational communication perspectives provide a framework for examining the impact of new Computerised Physician Order Entry (CPOE) systems on health care organisations. The aim of this study was to utilise performance metrics (volume of telephone/fax calls and the management of unfulfilled test requests) as a way of monitoring the impact of a new CPOE system on the communication (synchronous and asynchronous) interface in the Central Specimen Reception (CSR) area of a pathology laboratory service. The total number of outgoing and incoming calls rose considerably after the implementation of the new system. The number of unfulfilled test requests initially increased in the implementation period and thereafter fell to below pre-implementation levels. There were significant differences in the relative proportion of duplicate (69% - 35%) and rescheduled requests (4% - 24%) between the pre- and post-periods. Performance metrics, can be relevant for measuring and monitoring changes in communication processes. This is important with CPOE systems whose introduction can have unexpected consequences requiring early detection and action.

Keywords. Computer order entry, Evaluation studies, Hospital information systems, Laboratories, Pathology

Introduction

CPOE systems automate the clinical ordering process [1], and through the incorporation of clinical decision support and database linkage have the potential to contribute to improving the efficiency and quality of health care delivery [2]. However their introduction into hospitals can also result in major changes to work practices

¹ Andrew Georgiou, Health Informatics Research & Evaluation Unit, Faculty of Health Sciences, The University of Sydney, Cumberland Campus, East Street PO Box 170 Lidcombe NSW 1825 Australia. E-mail: a.georgiou@usyd.edu.au URL address: www.fhs.usyd.edu.au/hireu

particularly in the way that hospital departments communicate and work with each other [3]. To date there has been little research into the impact of these systems on laboratory functioning. The existing research in this area has tended to focus on the pre-analytical (doctor's decision to order) and post-analytical (delivery and application of test results) stages of the pathology ordering process, with little attention to ward-laboratory communication patterns and work patterns [4].

Information processing and communications are critical features of most activities within an organisation. Careful and systematic monitoring of how CPOE systems are used and their contribution to these processes can help to maximise system effectiveness [5]. The aim of this study was to utilise two performance metrics (volume of telephone; and fax calls and the management of unfulfilled test requests) as a means of monitoring the impact of a new CPOE system on the communication (synchronous and asynchronous) interface in the Central Specimen Reception (CSR) area of a pathology laboratory service after the changeover to a new results reporting system followed by electronic ordering two months later.

1. Material and Methods

1.1. Research Setting

The study was undertaken in the CSR department (consisting of around 20 staff) of a pathology laboratory service located in a large (640-bed) hospital in Sydney, Australia. On 22 November 2005, the Cerner Millennium Pathnet system replaced the previous laboratory information system. This was the precursor to the introduction in January 2006 of the Power Chart (version 2004.01) electronic ordering system which included some basic decision support features including prompts for essential patient information and notification of duplicate test requests.

1.2. Procedures

1.2.1. Telephone communications

Hospital communication data logs listing the number of incoming and outgoing calls for each of the existing CSR phones and fax machines were accessed. These summaries were grouped into five quarters beginning in June – August 2005 and ending June – August 2006 to compare the number of calls over the period.

1.2.2. Unfulfilled test requests

In the pre-CPOE period, CSR blood collectors visited wards to access the hand written requests. They matched the hand written request with the patient, and then proceeded with the specimen collection. On occasions where a collection was unable to be taken, and after consultation with the responsible clinician, the request was either put aside for collection on a future round or a notation was made on the request form and then returned to CSR as an unfulfilled request. *An unfulfilled request is therefore one where a blood specimen was not taken and the request cancelled.* Unfulfilled requests can occur for a number of reasons; it could be a duplicate test request inadvertently made for the same patient by different clinicians; it may have been cancelled by the clinician;

or it may have been reissued as a new request. Unfulfilled request forms were stored for an indefinite period before being discarded.

This procedure changed with the implementation of the new system on 22 November 2005. The department introduced a form to record the details of the episode including patient identification, ward, and date, and the reason for not collecting a specimen. The form provided the following choices: 1) Difficult collection; 2) Patient refused; 3) Patient unavailable; 4) Patient aggressive; 5) Patient not fasting; 6) Other. The forms also required the collectors to record whether the collection was rescheduled or cancelled. The information from these forms was then used to either cancel or reschedule the test request in the electronic system. These forms, along with all unfulfilled requests forms prior to 22 November 2005, were made available to the research team to audit for the period September 2005 to March 2006. Data were collated and cross validated by two researchers. Data about the total number of test requests per month were obtained from the pathology information service.

1.2.3. Analysis

The total and average number of incoming and outgoing calls per telephone/fax line were analysed by three-month (quarterly) periods. The proportion of unfulfilled requests to the total number of tests over each month, and of telephone calls for each quarter were also calculated. To aid the longitudinal overview of these data, the month of November (unfulfilled requests) and the Sep-Nov quarter (phone calls) were included as part of the pre-implementation period. However the Chi-square tests comparing types of unfulfilled requests during the pre- and post- periods used 22 November 2005 as the delineator date.

2. Results

2.1. Telephone communication

Table 1 shows the number of outgoing and incoming calls per quarter alongside their proportion to the total number of test requests for each period. It also provides the average number of calls per telephone/fax line. The number of calls (incoming and outgoing) for each quarter doubled in the Mar-May 06 periods and remained high in the Jun-Aug 06 period. The average number of incoming calls over the study period changed considerably but with high standard deviation (SD) values. In contrast the averages for outgoing calls did not vary as much and the SD values were lower.

2.2. Unfulfilled test requests

There were 4794 unfulfilled test requests for the period September 2005 to March 2006. Table 2 shows that the number of unfulfilled test requests rose sharply from 356 in the pre-implementation month of September 2005, to a peak of 1543 in December 2005, and then fell to 143 in March 2006. There was a similar trend in the proportion of unfulfilled test requests to total test requests, rising from 0.008 in September 2005 to 0.04 in December 2005 and then decreasing to 0.003 in March 2006. The number of cancelled and rescheduled requests was also compared over the pre- and post-implementation periods. In the pre-implementation period rescheduled requests

amounted to 4% (n=26) of all unfulfilled requests. This proportion rose to 24% (n=969) post-implementation. Cancelled requests fell from 96% (n=672) of the total pre-implementation number to 76% (n=3127) in the post-implementation period ($\chi^2 = 144.1$; df 1; $p < 0.0001$). There was also a significant decrease in the proportion of duplicate requests from 69% (n=484) to 35% (n=1448) ($\chi^2 = 286.4$; df 1; $p < 0.0001$).

Table 1 Total and mean (with SD) of calls per telephone/fax of incoming/outgoing phone calls and the proportion of total requests (pre-implementation quarters shaded)

	No.total requests	No. outgoing calls (Proportion to total requests)	Mean (SD)	No. incoming calls (Proportion to total requests)	Mean (SD)
Jun-Aug 05	121290	2037 (0.02)	169.8 (95.7)	1268 (0.01)	105.7 (56.2)
Sep-Nov 05	121372	2872 (0.02)	119.7 (68.2)	4054 (0.02)	168.9 (407.6)
Dec-Feb 06	111703	3061 (0.03)	145.8 (81.5)	4871 (0.04)	232.0 (477.6)
Mar-May 06	118290	6078 (0.05)	155.9 (96.1)	10683 (0.09)	273.9 (554.8)
Jun-Aug 06	125334	5850 (0.05)	121.9 (87.5)	10678 (0.09)	222.5 (490.1)

Table 2 No. unfulfilled requests as a proportion of total requests (pre-implementation months shaded)

Month	No. unfulfilled requests	No. total requests	Proportion
Sep-05	356	42066	0.008
Oct-05	323	39551	0.008
Nov-05	395	39755	0.010
Dec-05	1543	38129	0.040
Jan-06	1234	36559	0.034
Feb-06	800	37015	0.022
Mar-06	143	42513	0.003

3. Discussion

The results show dramatic fluctuations in the number of telephone calls and unfulfilled test requests from the period prior to the system changeover and extending some months later. These fluctuations can impact on the synchronous and asynchronous channels of communication with consequences for work processes in the department.

3.1. Synchronous communication

The results of the comparison of telephone calls revealed a major increase in the number of incoming and outgoing phone calls associated with the introduction of the new reporting system in November 2005 followed by the new ordering system in January 2006. This implies a rise in the level of activity within the department. The

high standard deviation values for the means of incoming calls suggests that the increased number of calls has not occurred in a sustained way across the department, and are possibly concentrated in those sections which deal with enquiries from the wards. Conversely, the smaller standard deviation values for outgoing calls indicate that the increased number of calls was more consistent across the department. These findings correspond with research suggesting that changes in modes of communication brought about by the introduction of asynchronous CPOE order channels can contribute to levels of disruption and dysfunction [6].

3.2. Asynchronous communication

In the pre-implementation period the recording of a reason for an unfulfilled test request was generally *ad hoc* and inconsistent. This procedure was standardised after the introduction of the new results reporting system on 22 November 2005. The introduction of structured information allows clinicians to electronically monitor the status of requests. It also indicates a higher level of CSR/ward accountability. However, the rise in the volume of telephone calls beginning with the introduction of the new results reporting system (November 2005) followed by the new order entry system (January 2006) suggests that the phone was used heavily by clinicians as a means of monitoring the status of test requests. This may have been a transitory phenomenon associated with unfamiliarity of the new system [7]. Regular monitoring of the situation using the metrics outlined in this study can answer this question.

While the proportion of unfulfilled requests increased dramatically following the system changeover on 22 November 2005, it fell away after a few months to levels below those found previously. This rise is possibly due to the instability associated with implementation. On the other hand, the significant decrease in the relative proportion of duplicate requests points to the existence of a more fundamental change associated with the new system. This supports existing evidence that CPOE can help to reduce the level of unnecessary and duplicate requests [8].

The fall in the number of cancelled requests as a proportion of all unfulfilled requests is more complicated. There are instances where it is obviously necessary to cancel a test request. Such an occasion occurs when a patient is discharged or a test request has been duplicated by mistake, or even when a doctor may decide to cancel a request. However, not all unfulfilled requests need to be cancelled. For instance, a patient may be temporarily unavailable or may not have fasted, or there may have been a situation where a collection was not possible. A patient may not be available for a blood collection for no other reason than they were undergoing treatment in another part of the hospital at the time. The decrease in the relative proportion of cancelled requests is therefore likely to be a consequence of the replacement of previous *ad hoc* monitoring systems with improved reporting structures associated with CPOE.

3.3. Laboratory impact

CSR occupies a specific organisational role in the laboratory test process sitting between the clinician's decision to make a test request and the actual processing of the specimen [3]. Its responsibilities include the maintenance of maximum levels of *coordination* (of information and specimens), as the preservation of the *integrity* of the test request. This in turn involves attention to *accuracy* and requires high levels of *accountability* and *efficiency*. The results of this study show that CPOE can impact on

these areas of responsibility. This can occur through the introduction of structured ways of entering data which can lead to improved levels of coordination and accountability. It can also lead to changes in the efficiency of work processes, especially through its ability to reduce duplication. However, these changes are not necessarily consistent. The increased levels of telephone/fax communication in the department associated with the system changeover suggest that it may also severely affect work load levels.

3.4. Limitations:

The choice of research method, in this case the monitoring of telephone/fax communications and unfulfilled orders can be affected by issues of data comprehensiveness and reliability. This study has endeavoured to offset these potential limitations through rigorous attention to the accuracy and completeness of the data.

4. Conclusion

Communication within the hospital setting is all pervasive but is often overlooked or taken for granted. Performance metrics, chosen wisely and used carefully, can be relevant to the task of monitoring changes in communication processes. They can also serve as a valuable tool for identifying trends or potential problems as part of statistical process control methods aimed at the early detection and prevention of problems [9]. This is particularly important with CPOE systems where implementation can have unexpected (possibly dysfunctional) consequences requiring early detection and action.

Acknowledgements

The authors acknowledge the cooperation and support of CSR and hospital staff in this study. The research was conducted as part of an Australian Research Council Linkage Grant in collaboration with the New South Wales Health Department.

References

- [1] Kuperman, G.J. and R.F. Gibson, *Computer physician order entry: benefits, costs, and issues*. Ann Intern Med, 2003. **139**(1): p. 31-9.
- [2] Abelson, D., D. Connelly, G. Klee, et al., *Technology Assessment: Computerized Physician Order Entry*. 2001, Institute for Clinical Systems Improvement.
- [3] Georgiou, A., J.I. Westbrook, J. Braithwaite, et al., *When requests become orders - A formative investigation into the impact of a computerized physician order entry system on a pathology laboratory service*. Int J Med Inf, 2007. **76**(8): p. 583-591.
- [4] Georgiou, A., M. Williamson, J. Westbrook, et al., *The impact of computerised physician order entry systems on pathology services: a systematic review*. Int J Med Inf, 2007. **76**(7): p. 514-29.
- [5] Beuscart-Zephir, M.C., S. Pelayo, P. Degoulet, et al., *A usability study of CPOE's medication administration functions: impact on physician-nurse cooperation*. Medinfo., 2004. **11**(Pt 2): p. 1018-22.
- [6] Beuscart-Zephir, M.C., S. Pelayo, F. Anceaux, et al., *Impact of CPOE on doctor-nurse cooperation for the medication ordering and administration process*. Int J Med Inf, 2005. **74**(7-8): p. 629-41.
- [7] Davidson, E.J. and W. Chismar, *Examining the organizational implications of IT use in hospital-based health care: a case study of computerized order entry*. Proceedings of the 32nd Hawaii International Conference on System Sciences, 1999.
- [8] Bates, D., G. Kuperman, M. Rittenberg, et al., *A randomized trial of a computer-based intervention to reduce utilization of redundant laboratory tests*. Am J Med, 1999. **106** (February): p. 144-150.
- [9] Smith, G., *Statistical process control and quality improvement*. 1991, New York: Merrill.

Development of a patient-oriented tool for evaluating the quality of breast cancer information on the Internet

HSU Wen-Chin¹, BATH Peter A.

Health Informatics Research Group, Department of Information Studies, University of Sheffield, Sheffield, UK.

Abstract. The aim of this study was to develop a tool for evaluating the quality of breast cancer information on the Internet from the perspective of patients and their families. A specific tool, Breast Cancer tool (BC tool), was developed based on the information needs of women with breast cancer and their families reported in the literature. The BC tool and other 3 generic tools (HON, IQ tool, Discern) were used to assess 40 breast cancer websites. The reliability and validity of each tool was examined and the time spent reviewing the websites was measured. The four tools were shown to have acceptable reliability (Cronbach's $\alpha > 0.7$), convergent validity, especially the BC tool which was capable of distinguishing whether a website offers sufficient information for women and their families. However, the BC tool took more time than the other tools to use, suggesting relatively low feasibility. The results of this study reinforce the importance of developing specific tools from perspectives of patients and their family members.

Keywords. Evaluation tool, Reliability, Validity, Feasibility, Information quality

1. Introduction

The importance of developing a patient-orientated tool for evaluating medical information on the Internet has been highlighted [1][2]. Although numerous tools have been developed to guide the layperson away from potentially harmful websites [3], these tools are usually designed to assess generic types of health information using pre-defined quality criteria. Several criticisms have been directed at this kind of tool: first, even though websites score well using the quality criteria, the content may be quite poor [4]; second, most of these evaluation tools are designed from the perspective of health professionals, rather than the patients and their family members; third, there is a lack of consensus about quality of health information on the Internet; fourth, although many organizations have established quality criteria to guide and assess health-related

¹ Corresponding Author: Wen-Chin Hsu, Health Informatics Research Group, Department of Information Studies, University of Sheffield, Sheffield, UK. Email: lip06wh@sheffield.ac.uk.

website content (e.g. HON Code, American Medical Association), the reliability and validity of these evaluation tools are unknown.

To overcome these problems, several studies (e.g. Bouchier and Bath, 2003 [5]) have suggested that there is a need to develop specific evaluation tools that assess whether the information meets the needs of particular groups, e.g., for patients/carers affected by multiple sclerosis [2], Alzheimer's disease [5], and diabetes [6]. No tool was found that had been developed for breast cancer, the most common cause of cancer deaths in women worldwide [7]. Relatively few studies have tested the reliability and validity of their evaluation instruments [8]. The aim of this study was to develop and test a tool for evaluating the quality of breast cancer information on the Internet based on the reported information needs of patients and their families.

2. Methods

The purpose of the Breast Cancer (BC) tool was to determine whether the information provided on a website meets the needs of patients and families. To establish the evaluation criteria of the BC tool, a comprehensive search was conducted using five medical databases (Medline, EMBASE, PsycInfo, CINAHL, Web of Knowledge) using the search terms '*(information needs) AND (breast cancer) AND (family OR husband OR partner OR women OR spouse)*'.for articles published in English between January 1996 and May 2006. This period (1996-2006) was chosen because a previous review [9] had identified the information needs of women with breast cancer and their family members published between 1988 and 1996. The following studies were included: review-type articles focusing on the information needs of, or the information sources and providers for, women with breast cancer and/or their family members. The search resulted in a total of 369 articles, and each article was examined. Twenty-one articles related to the focus of the study were identified and further classified into several categories to help identify whether the target websites provides sufficient information for people at different stages of breast cancer. These categories included information about the disease, symptoms and diagnosis, treatment and breast re-construction, recovery, etc. The items in these categories were then formulated into 47 questions for assessing whether a web-site contained this information. An example of one question is 'Does the site provide information about the stage of breast cancer?' For each question a three-point scale (Yes, Partly/Not sure, No) was used to score the extent to which information on that specific issue was present.

Three generic tools (HON code [10], Information Quality (IQ) Tool [11] and Discern [12]) were selected to compare their reliability, validity and feasibility with those of the BC tool because they were developed by authoritative organizations and have satisfactory internal consistency and inter-rater reliability (IQ Tool and Discern) or been successfully applied by previous researchers (HON code) [2][13][14]. The characteristics of these tools are shown in Table 1.

A search of breast cancer websites was conducted on using three search engines with the term 'breast cancer'. The top 100 web sites identified from each search engine were reviewed, resulting in an initial sample of 300 websites, of which 260 sites (86%) were excluded because they were either broken links, duplicated websites, provided

little or no breast cancer information, or a fee or registration was required for accessing the information. A final sample of 40 breast cancer websites was included for analysis.

Table 1: Characteristics of the tools used in the study

Title of the tool	Developer	No. of questions	Quality Measures
HON Code Site Checker	Health on the Net Foundation (2006)	17*	Authority, Purpose of the website, Privacy, Referenced and dated, Justification of claims, Website contact details, Disclosure of funding sources, Advertising policy
Information Quality (IQ) Tool	Health Summit Working Group (1999)	21	Authorship, sponsorship, currency, accuracy, confidentiality, navigability
DISCERN	University of Oxford and the British Library (1997)	16	Reliability, objectivity, information quality of treatment choice
Breast Cancer tool (BC tool)	Authors in this study	47	Information needs of women with breast cancer and their families

* The HON Code consisted of 14 questions. However, some of these questions are open-ended question (e.g. What/How), which are not easily applied to assessing websites. Therefore, considering the convenience of evaluation, the original version of the HON Code was rewritten as 17 closed-ended questions (e.g. Do/Are/Is)

Each website was evaluated in sequence using the three general and one specific tools tool (BC tool) by one of the authors (Hsu). The time spent on each evaluation tool was recorded in order to test the feasibility of using the tools, i.e., whether they were too long for individuals to use them feasibly. Data were inputted into Microsoft Office™ Excel 2003 and imported into Statistical Package for the Social Sciences™ (SPSS) version 13.0 for statistical analyses. Cronbach's Alpha Coefficient was used to test the internal reliability of each tool. Kendall's Coefficient of Concordance and Kendall's Tau B were chosen to test the convergent validity of the four evaluation tools. The time spent to review each web-site across the four tools was compared.

3. Results

The time taken to assess the websites was: IQ tool (mean = 6.43 minutes; SD=2.62); HON Code (mean = 8.24; range = 1.35-28.12; SD= 4.94); DISCERN (mean = 7.36; range = 3.55-16.51; SD = 2.60); BC tool (mean = 14.98; range = 8.10-34.00; SD = 5.41). There was a significant difference among the four tools in the time taken to review the web-sites ($p < 0.05$). The internal consistency of the 4 tools was as follows: IQ tool (Cronbach's alpha = 0.766); HON code (0.817); Discern (0.816); BC tool (0.876). The agreement among the tools for ranking the web-sites was significant ($p < 0.001$), but the degree of concordance was fair (Kendall's $w = 0.426$). The concordance among the tools reduced to $w = 0.301$ when the BC tool was excluded. Kendall's Tau B was used to examine further the agreement between each tool. The correlations between all of the tools was significant ($p < 0.05$), although there were no strong correlations (i.e., > 0.7). The strongest correlation was between the HON and IQ tool at 0.504. There was moderate correlation between the following tools : Discern and HON (0.458), Discern and IQ tool (0.448), Discern and BC tool (0.430). The

weakest correlations were between the BC tool and the HON (0.255), and the BC tool and the IQ tool (0.238).

4. Discussion

Feasibility refers to how long it takes to examine the websites, and whether the examiner had difficulty in using the tools [6]. In this study, the time spent reviewing the websites using each tool was measured to test for differences in the mean time spent reviewing the websites. The observed differences, particularly for the longer time taken using the BC tool, were probably because the BC tool contained considerably more questions than the other three tools, and because the four tools respectively examine different aspects of quality of the websites. The BC tool aims to examine whether the content of the websites meet the needs of women with breast cancer and their families while the others are designed to evaluate generic quality criteria, such as accuracy and objectivity. Examining some criteria can be time-consuming, e.g., the question in the BC tool: “Does the site provide information about the prevention of the treatment’s side effects?”. In some cases, a web-site only provided information about the treatment’s side effects and nothing about their prevention, and consequently, nothing was found to relate to the question after a lot of time had been spent reviewing the website. Compared with the 3 generic tools, the criteria in the BC tool were also harder to assess, being more subjective in nature.

Although the internal consistency of each tool was satisfactory (Cronbach’s $\alpha > 0.7$) in this study, meeting the requirements for acceptable reliability suggested by Bland and Altman (1997) [15], there were differences in the values of Cronbach’s α for the tools used in this study and the results reported in previous studies, shown in Table 2. For example, the reliability of the HON code tool was higher in this study (0.817) than in Harland and Bath’s study of web-sites concerning multiple sclerosis (0.537). The reliability of the IQ tool however, was lower in this study (0.766) than in Harland and Bath’s study (0.842), but higher than in Ademiluyi et al.’s study of smoking cessation information (0.634). The reliability of the Discern tool was higher in this study (0.816) than in Ademiluyi et al.’s study (0.777). The reliability of the BC tool (0.876) was lower than that of the MS tool developed by Harland and Bath (0.930) [2].

Table 2: Comparison of Cronbach’s α of different studies

Evaluation tools used	Cronbach’s α		
	This study	Ademiluyi et al. (2003) [14]	Harland and Bath (2007) [2]
HON code	0.817	-	0.537
IQ tool	0.766	0.634	0.842
Discern	0.816	0.777	-
BC tool	0.876	-	-
MS tool	-	-	0.930

These comparisons are relevant because this confirmed that the result of this study is reliable. That is, IQ tool and Discern were separately used in previous studies [2,14], and had shown to have acceptable reliability (Cronbach's $\alpha > 0.7$).

There are several possible explanations for these differences. First, the selected websites were different. This study focuses on breast cancer websites while Ademiluyi et al. [14] and Harland and Bath [2] separately concentrate on smoking cessation information and Multiple Sclerosis websites respectively. Second, the numbers of sampled websites was different. Ademiluyi et al. [14] sampled 89 websites while Harland and Bath [2] and this study sampled 40 websites. All of these samples constitute less than 100 cases, the minimum number which Kline (2000) [16] suggested to be essential to minimize the standard error of reliability statistics. Moreover, Ademiluyi et al. [14] also concluded that the sample size used in their study may be too small to ascertain the relationships between the scales' total scores, leading to a Type II error. The numbers of questions varies. The HON code employed by Harland and Bath [2] consisted of 9 questions, whereas the HON code used in this study was formulated into 14 questions. A greater number of questions are likely to increase the reliability if the tool is internally consistent. Third, the evaluation procedures were dissimilar. Two examiners were involved in the evaluation produce conducted by Ademiluyi et al. [14] while there was only one examiner participating in the evaluation produce conducted by Harland and Bath [2], and in this study. Furthermore, the sequence of using the tools to assess the websites may also affect the results. Fourth, the methods of recording the answers are different. In this study, the answers recorded in the HON code are divided into three categories (Yes/Not sure/Partly/No). However, a five-point-Likert-scale is used in Discern.

Overall, all of the four tools measure different quality concepts. The HON and IQ tools measure more similar concepts of quality than the other tools. The quality concepts that the BC tools measure appear to be very different from those of the other tools. This is not surprising because, basically, the BC tool measures different concepts of quality, regarding the information needs of women and their families, rather than the generic quality criteria.

Several limitations must be taken into account when explaining the results in this study. First, because only one reviewer evaluated all of the breast cancer websites, it is not possible to draw definite conclusions about its suitability as a quality assurance tool. Second, because of the dynamic nature of the Internet the results produced in this study may not be generalized to current population of breast cancer web-sites. Third, while the BC tool was designed to determine how well the web-sites met the cognitive needs of women with breast cancer and their families, e.g., information for decision-making, they do not give a clear indication of how well they might meet their affective needs, e.g., for emotional support. Finally, the BC tool was designed to assess the content of information on web-sites, and while it may be useful as a guide for health care professionals to evaluate web-sites before recommending them to individuals, it may be less useful for individuals themselves to use.

Due to time and financial restrictions, the BC tool was only used to assess the breast cancer websites once, and was not refined in the light of the results. Future researchers could modify this tool by using factor analysis, to identify the most important factors contributing to information content, and thus increase its feasibility for regular use.

5. Conclusion

The BC tool had good reliability and demonstrated some degree of convergent validity with other generic tools to assess the content of web-sites. The tool provides a means to evaluate how well web-sites meet the information needs of women with breast cancer and their families based on previous studies of information need in this group. Evaluating web-sites using tools customized for specific diseases complements the use of more generic tools and organizations such as MedCIRCLE, which evaluate and certify health information on the web. More generally, this study shows how a tool for evaluating the quality of information provided on the websites can be developed from the research literature. In practice, the breast cancer tool can be used by breast cancer organizations and support groups who wish to evaluate web-sites before providing access to them for their members. The tool could also be used by organisations who wish to improve the quality of websites providing information specifically relevant to breast cancer. This approach could be applied to develop tools for other diseases.

6. References

- [1] Coleman, B. Producing an information leaflet to help patients access high quality drug information on the Internet: a local study. *Health Information and Libraries Journal*. 2003;20:160-171.
- [2] Harland J, Bath PA. Assessing the quality of websites providing information on multiple sclerosis: evaluating tools and comparing sites. *Health Informatics Journal*. 2007 ;13 (3):207-221
- [3] Kim P, Thomas RE, Deering MJ, Maxfield A. Published criteria for evaluating health related web sites: review. *BMJ*. 1999;318 (7184):647-649.
- [4] Martin-Facklam M, Kostrzwa M, Schubert F, Gasse C, Haefeli WE. Quality markers of drug information on the internet: an evaluation of sites about St. John's Wort. *The American Journal of Medicine*. 2002;113:740-745.
- [5] Bouchier H, Bath PA. Evaluation of websites that provide information on Alzheimer's disease. *Health Informatics Journal*. 2003;9(1):17-31.
- [6] Seidman JJ, Steinwachs D, Rubin HR. Design and testing of a tool for evaluating the quality of diabetes consumer-information web sites. *Journal of Medical Internet Research [Online]*. 2003;5(4). [Accessed 24 July 2006] <http://www.jmir.org/2003/4/e30/index.htm>
- [7] Parkin DM, Bray F, Ferlay J, Pisani P. *Global Cancer Statistics 2002*. CA: A Cancer Journal for Clinicians. 2005;55:74-108.
- [8] Eysenbach G, Kohler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *British Medical Journal*. 2002;324:573-576.
- [9] Rees CE, Bath PA. The information needs and source preference of women with breast cancer and their family members: a review of the literature published between 1988 and 1998, *Journal of Advanced Nursing*. 2000;31:833-841.
- [10] Brooks R. The reliability and validity of the Health of the Nation Outcome Scales: validation in relation to patient derived measures. *Australian and New Zealand Journal of Psychiatry*. 2000;34(3):504-11.
- [11] IQ Tool <http://hitiweb.mitrectek.org/iq/iqmain.asp>
- [12] DISCERN background, general instructions and instrument (1997). Available from: <http://www.discern.org.uk>
- [13] Rees CE, Ford JE, Sheard CE. Evaluating the reliability of DISCERN: a tool for assessing the quality of written patient info. on treatment choices. *Patient education and counseling*. 2002;47(3):273-275.
- [14] Ademiluyi G, Rees CE, Sheard CE. Evaluating the reliability and validity of three tools to assess the quality of health information on the Internet. *Patient Educ Couns*. 2003;50(2):151-155.
- [15] Bland JM, Altman DG. Cronbach's alpha. *Br Med J*. 1997;314:571
- [16] Kline P. *The handbook of psychological testing*. London: Routledge. 2000.

Pre-Post Evaluation of Physicians' Satisfaction with a Redesigned Electronic Medical Record System

Monique W.M. JASPERS^a, Linda W.P. PEUTE ^a, Arnaud LAUTESLAGER^b, Piet J. M. BAKKER^c

^a *Department of Medical Informatics, Academic Medical Center, University of Amsterdam, the Netherlands*

^b *ICT department, Academic Medical center, Amsterdam,*

^c *Department of Medical Oncology, Academic Medical Center, Amsterdam*

Abstract: Physicians' acceptance of Electronic Medical Record Systems (EMRs) is closely related to their usability. Knowledge about end-users' opinions on usability of an EMR system may contribute to planning for the next phase of the usability cycle of the system. A demand for integration of new functionalities, such as computerized order entry and an electronic patient status led to redesign of our EMR system, which had been in use for over 8 years at the Academic Medical Center of Amsterdam. The aim of this study was to understand whether the redesigned EMR system was an improvement of the earlier EMR and which system aspects accounted for user satisfaction and which did not. We conducted a formative pre- and post usability evaluation of our former and redesigned EMR system. For the assessment of both system versions' usability, we distributed two standardized usability questionnaires among 150 clinicians who routinely had used the older EMR system and had been working with its newer version for 6 weeks. Though overall user satisfaction was relatively high for both EMR systems, screen layout and interaction structure proved less easy to work with in the newer EMR system. The new EMR system however was more appreciated because of its enhanced functionality, capabilities and likeable user-interface. The results point to a number of actions that might be useful in future usability improvement efforts of our EMR system and other EMRs.

Keywords: User Interfaces, EPR-CPR-EM, HER, Assessment-Evaluation, Health professional workstation

Introduction

Electronic Medical Records (EMRs) have the potential to advance the quality of care, patient safety, to facilitate work flow, decrease medical errors and reduce costs [1], and to improve communication among physicians [2,3]. The integration of order-entry systems and decision support further enhance their use [4,5]. At the same time, EMRs have disseminated slowly and many EMRs have failed their expectations, whereas health care organizations and their health care professionals nowadays make high

demands upon EMRs [6]. One of the largest barriers to EMR adoption has been resistance from physicians [7]. Factors influencing physicians' adoption of and satisfaction with EMRs have been system's ease of use, learnability and memorability, error frequency, system response time, and additional work load for physicians [8]. End users' adoption of and satisfaction with interactive computer systems indeed is directly influenced by user interface characteristics such as the extent to which the system functions are obvious and accessible to the user, labels and names are distinct from one another and follow the users' terminology and language, recognition is promoted instead of recall of information, and the natural workflow of the user is supported [9]. But even nowadays, design and evaluation of EMRs interfaces have not received much attention as a research area, though from the users' perspective the user interface is the only visible part of an EMR. If our aim is to increase physicians' acceptance of EMRs, we should know which features make EMRs usable and which should be improved.

This study was undertaken to measure usability and user satisfaction with an EMR in routine clinical use in a large academic medical center. In a pre-and post-evaluation, we compared physicians' user interaction satisfaction with the EMR before and after its redesign. The aim was to understand whether the redesigned EMR system was an improvement of the earlier EMR and which system aspects accounted for user satisfaction and which did not. This paper provides the results of this usability evaluation study. We conclude with lessons learned about end-users' satisfaction with EMR design aspects and about the next phase in the usability life cycle of our EMR.

1. Background setting

This study was performed within the Academic Medical Center (AMC), a large hospital organization in Amsterdam, The Netherlands. The AMC is a 1002 beds hospital including 21 outpatient clinics, 34 inpatient clinics, 5 day care units, employing 2200 clinicians in total. The development of an institute-wide integrated results review EMR was started in 1998; the 'PoliPlus' system. At that moment, the 'PoliPlus system' was an in-house developed web-based system. Its technology was grounded on the three-tier architecture. Easy accessibility and system simplicity were the system's primary design principles. For over 8 years, this first EMR system version provided financial, administrative, clinical results review and patient list management facilities.

The vast majority of clinicians had used this PoliPlus system in support of their daily-patient care practices. Due to a demand for increase of its functionality, the PoliPlus system was integrated with commercially based facilities for ordering medications and clinical tests and an electronic patient status. On the front-end, this functional enhancement of the EMR required redesign of the navigational structure. The provision of all old and new functionalities in a logical and consistent manner was required in the new interface, and the proven user friendliness of the EMR system was to be preserved. Due to vast use of Microsoft windows, the new system version integrated Microsoft Windows menu-based components to realize this aim. This new version of the EMR system, the 'AMC Care desktop' was introduced hospital-wide in 2006. Figure 1 displays an overview of the main screens of the former PoliPlus system (left) and the new EMR system, the AMC Care desktop (right). Clinician users were informed about the introduction and capabilities of the new EMR system by flyers

explaining the new system functions and providing a step through scenario for learning to use the new system. Besides, a digital demonstration video was available and clinicians were invited for demonstrations of the new system at their departments.

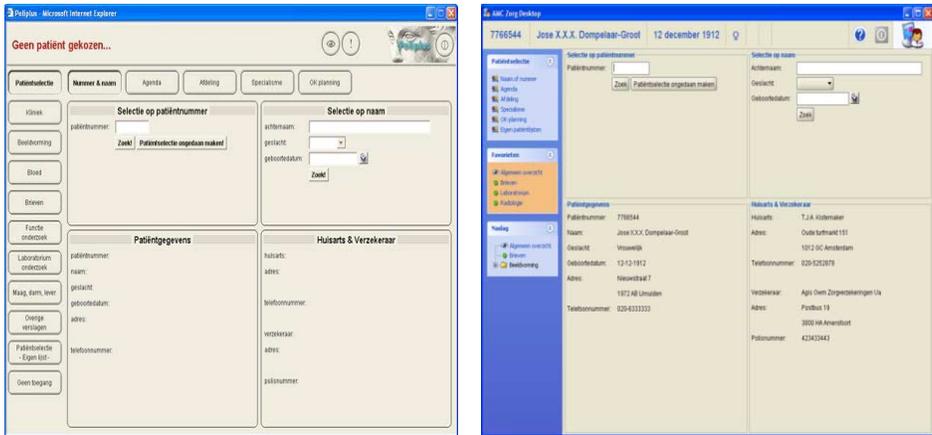


Figure 1. Screenshots of the old PoliPlus system (left) and the new EMR system (right) to offer a visual overview of the changed screen and navigation layout, from button structure (left) to window based (right).

2. Methods

We applied two standardized questionnaires to measure user satisfaction with the EMR system before and after its redesign: the Questionnaire for User Interaction Satisfaction (QUIS, see [10]), and the IBM Computer Usability Satisfaction Questionnaire [11]. The short-form QUIS was used to assess the EMR interface as regards overall user reaction to the system, user satisfaction as regards screen lay-out, terminology used, system learnability and system capabilities. More specifically, we were interested to know to what extent screen layout and navigation structure of the new EMR system supported the clinicians in searching for patient information. The IBM Computer Usability Satisfaction Questionnaire was used in addition to the QUIS to assess to what extent the system as compared to its earlier version was efficient, effective, responsive, learnable, and informative in its help functions. Both in the pre-and post evaluation phase of the EMR system, a group of 150 clinicians were asked to participate in the study via an interdepartmental e-mail explaining the purpose of the study, providing a web link for filling in the QUIS and IBM Computer Usability Satisfaction questionnaires. Additional validation and confidence of findings were pursued through a comparison of the quantitative questionnaires' results by a pre- and post-survey among the respondents allowing them to provide additional comments and remarks with regard to the older and newer version of the EMR system.

3. Results

In the pretest condition, 54 clinicians and in the posttest condition 62 clinicians filled out the questionnaires (36% and 41% respectively). Overlap in system respondents in the pre- and posttest was estimated on 80%. The pretest was organized three weeks before introduction of the new system. The posttest evaluation was organized six weeks after the introduction of the newer EMR system to give system users familiarity with the new system in daily clinical practice. The respondents covered all clinical departments within the AMC. The respondents of the pretest and posttest did not significantly differ in the distributions of age, general computer experience, and experience with the former PoliPlus system. The average age of the respondents was 44 years, computer experience ranged from intermediate (45%) to experienced (55%). Eleven percent of the respondents had 3-6 months experience, 18% had 6-11 months of experience, 32 % had 1-3 years of experience, and 39% had over 3 years experience with the former PoliPlus system. Forty-four percent of the respondents had used the EMR system 2-10 hours per week, 28% used the EMR system 10-20 hours per week and 15% used it over 20 hours per week. Eleven percent of the respondents had seen a demonstration of the new EMR system after which they immediately had started to use the new EMR system. Eighty-nine percent of the respondents had read the flyer introducing the EMR system, of which 44 % of the respondents had learned to use the new system by stepping through the scenario described in the flyer and 45% had started using the system without stepping through the scenario. Ninety-five percent of all respondents were of the opinion that no further training or instruction was required for learning to work with the new EMR system.

Table 1 shows the mean user response for the QUIS questionnaire in the pretest and posttest evaluation of the EMR system. The mean user responses for each QUIS section in the pre-and posttest condition were relatively high, ranging from 7.0 to 7.8. The mean user ratings of the categories easy, stimulating and flexible of the QUIS section 'overall user reaction' were significantly higher for the redesigned EMR compared to its earlier version. The newer EMR however scored lower on the question related to the organization of information on the screen QUIS section 'screen design and layout' than the older version. Remembering names and use of commands was somewhat more difficult, but error messages were more helpful and correcting mistakes was less complicated with the new EMR system. The IBM Computer Usability questionnaire shed some light on the satisfaction of the users with the newer EMR with regard to its efficiency, efficacy, responsiveness, learnability as compared to its previous version. Overall, users were as much satisfied with the new EMR system as with the older EMR system with regard to these aspects, but significantly more satisfied with the interface and with the functions and capabilities provided by the newer EMR system. These findings were confirmed by the survey results. Clinicians reported that the older EMR system was easy to work with because all patient information could be accessed and all system functionalities could be activated from the main screen. In the post-test survey, clinicians remarked that the new EMR system was an improvement because of its enhanced functionality, capabilities and likeable user interface, but was somewhat more difficult to work with because of the reorganization of information and functions on these screens.

Table 1. Results of the QUIS questionnaire for assessing user interface satisfaction with the older EMR compared to the new version EMR on a LIKERT scale of 1-9, with 1 as lowest and 9 as highest rating.

QUIS Section	Older EMR		New EMR	
	Mean user response	S.D	Mean user response	S.D.
Overall reaction to the system	7.1	0.4	7.6	0.2
Screen design and layout	7.8	0.2	7.0	0.2
Terminology and system information	7.5	0.4	7.4	0.3
System Learnability	8.1	0.2	7.6	0.2
System capabilities	6.7	0.3	7.0	0.2

4. Discussion/ conclusion

The pretest results suggest that prior to installation of the newer EMR system, clinicians overall had positive experiences with the older PoliPlus system, specifically with regard to its screen design and layout and the ease with which system commands could be memorized. Whereas the newer version of the EMR system proved to be more flexible and stimulating to its users, the organization of information on the screens proved to be less usable to its users than the former EMR version. These findings were not too surprising for two reasons. First, clinicians had been using the older PoliPlus system for at least 1 year and were thus rather proficient in operating the older EMR system. The introduction of the new EMR system, extended with order-entry ordering facilities and an electronic patient status brought with it a radically changed interaction structure and screen layout. The lower post test ratings of screen layout and ease of commands use in the new EMR system might be attributed to the short time period of six weeks in which clinicians had to become acquainted with the new EMR system. This fact may also account for clinicians' higher appreciation of error messages of and the way mistakes could be corrected in the new system. In learning to use the redesigned EMR system with its more sophisticated interaction structure, clinicians were probably more dependent upon these messages than in daily use of the older system, with which they had become very much familiar with. Despite the difficulties clinicians seemed to experience in learning to operate the new EMR system, they apparently appreciated it for its extended functionalities and capabilities, and even liked using the new interface more than the older user interface. What remains one of the greatest challenges of designing our EMR system user interface is balancing the clinicians' requirements with regard to an integrated fully-functional EMR and simple screen design and navigational structure. In many instances, navigation and orientation problems in the CPR make it hard for clinicians to gain a rapid overview of the patient's clinical problems [2, 12]. Too much information and functions on a single screen may make these hard to find and problematic when trying to perform tasks efficiently. Less information and functions on a single screen will require more screens to provide these all and thus will lead to a more complex user-interaction structure. Minor changes in screen designs and interaction structure can have a major impact on the usability of a system. The context and interaction issues discovered in this study can be linked back to the importance of user-centred system design. For the interaction structure being effective, the tasks and procedures that the user may perform with an

EMR system need to be structured in a logical and consistent manner. This means that the system functions should correspond one to one with the goals that the user sets himself in performing his tasks and the order in which the user wants to attain these goals. Also, the presentation order of information should match the order in which a user processes this information. And as a general rule it can be stated that information that has to be memorized is more easily remembered if data elements that are related are clustered on the computer screen [13].

A further usability improvement of the EMR system will be necessary to determine how the clinicians' needs can be even more accommodated than with our current version. The results of our questionnaire contribute to extensive discussions for modification of the system on the usability issues revealed. Understanding clinicians' activities, information needs in the context of performing these activities and communication patterns is crucial in designing a user interface that is truly usable.

The next phase in the usability life cycle of our EMR is to conduct extensive cognitive analysis studies of clinicians' daily activities, types of information needed and communication patterns. We recently conducted a study focused on a detailed analysis of the physician-nurse communication patterns and information exchanges during the morning report. The results of this study and others will be used to redesign our EMR computerized record system that will support our clinicians in performing their tasks supported by the EMR system in an even more user-friendly way.

References

- [1] S.J.Wang, Middleton B, Prosser LA, Bardon CG, Spurr CD, Carchidi PJ, et al. A cost-benefit analysis of electronic medical records in primary care. *Am J Med.* 2003; 114: 397-403.
- [2] Hier DB, Rothschild A, LeMaistre A, Keeler J. Differing faculty and housestaff acceptance of an electronic health record. *Int J Med Inform.* 2005; 74: 657-662.
- [3] Embi PJ, Yackel TR, Logan JR, Bowen JL, Cooney TG, Gorman PN. Impacts of computerized physician documentation in a teaching hospital: perceptions of faculty and resident physicians. *J Am Med Inform Assoc.* 2004; 11: 300-309.
- [4] Bates DW, Kuperman GJ, Wang S, Gandhi T, Kittler A, Volk L, et al. Ten Commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. *J Am Med Inform Assoc.* 2003; 10: 523-530.
- [5] Poissant L, Pereira J, Tamblyn R, Kawasumi Y. The impact of electronic health care records on time efficiency of physicians and nurses: a systematic review. *J Am Med Inform Assoc.* 2005; 12: 505-516
- [6] Jaspers MWM, Knaup P, Schmidt D. The computerized record system: where do we stand? *Yearbook of Medical Informatics 2006: Methods of Information in Medicine 2006: Suppl 1:29-39.*
- [7] Rose AF, Schnipper JL, Park ER, Poon EG, Li Q, Middleton B. Using qualitative studies to improve the usability of an EMR. *Journal of Biomedical Informatics* 2005; 38: 51-60.
- [8] Delpierre C, Cuzin L, Fillaux J, Alvarez M, Massip P, Lang T. A systematic review of computer-based patient record systems and quality of care: more clinical trials or a broader approach? *Int J Qual Health Care* 2004; 16: 407-416.
- [9] Nielsen J, *Usability Engineering*, Academic Press, Toronto, 1993.
- [10] Chin JP, Diehl VA, Norman KA. Development of an instrument measuring user satisfaction of the human-computer interface. *ACM CHI'88 Proceedings* (1988) 213-218.
- [11] Lewis JR. IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction* 1995; 7:1: 57-78.
- [12] Jaspers MWM, Steen T, Van den Bos C, Geenen M. The think-aloud method: a guide to user interface design. *Int J Med Inform.* 2004; 73: 781-795.
- [13] Gardiner MM, Christie B. *Applying cognitive psychology to user interface design*, John Wiley & Sons, Chichester, 1994.

CPOE System Design Aspects and Their Qualitative Effect on Usability

Reza KHAJOUEI^{a,b,1}, Monique W. M. JASPERS^a

^a*Department of Medical Informatics, Academic Medical Center-
University of Amsterdam, The Netherlands*

^b*Kerman University of Medical Sciences, Iran*

Abstract: Although many studies have discussed the benefits of Computerized Provider Order Entry (CPOE) systems, their configuration can have a great impact on clinicians' adoption of these systems. Poorly designed CPOE systems can lead to usability problems, users' dissatisfaction and may disrupt normal flow of clinical activities. This paper reports on a literature review focused on the identification of CPOE medication systems' design aspects that impact CPOE systems' usability and create opportunities for medication errors. Our review is based on a systematic literature search in PubMed, EMBASE and Ovid MEDLINE for relevant publications from 1986-2006. We categorized the design aspects extracted from relevant publications into six different groups: 1) timing of alerts, 2) log in/out procedures, 3) pick lists and drop down menus, 4) clues and guidelines, 5) documentation and data entry options, and 6) screen display and layout. Our review shows that the manner in which a CPOE system is configured can have a high impact on ease of system use, task behavior of clinicians in ordering drugs, and medication errors. Characterization of consequences associated with certain CPOE design aspects provides insight into how CPOE system designs can be improved to enhance physicians' adoption of these systems and their success. Recommendations are provided to enable CPOE system designers to create CPOE systems that are not only more user friendly and efficient but safer.

Keywords: CPOE, medication systems, medication errors, assessment-evaluation, design aspects, user-computer interface, usability evaluation, human factors

Introduction

Computerized Provider Order Entry (CPOE) systems can have a significant impact on the safety and quality of drug management [1;2]. The configuration of CPOE yet affects physicians' adoption and usage of these systems, making these systems often difficult to learn and complicated to use. CPOE interface design often does not conform to clinicians' decision-making and workflow processes making the

¹ Corresponding author: Department of Medical Informatics, Academic Medical Center, PO Box 22700, 1105 AZ Amsterdam, The Netherlands, E-mail: r.khajouei@amc.uva.nl

appropriate order entry strategy not apparent to CPOE users[3]. Poor CPOE interface design and lack of usability facilitates medical error and may even lead to disaster if critical information is not presented in an effective manner. Both quantitative and qualitative studies have highlighted examples of CPOE system design flaws that indeed lead to errors in orders. Many adverse drug events for example result from poor interface design rather than from human error [4-6]. Complex CPOE systems place heavy cognitive demands on the users and may result in suboptimal use of features designed to support clinicians in the ordering process[7;8]. Use of effective external representations could yet facilitate the completion of order entry task[7;8]. However, despite the great impact of CPOE design on usability and medical errors, no literature reviews have focused specifically on the influence that certain CPOE design aspects can have on the difficulties that clinicians encounter when using a CPOE system. The purpose of this study was to review current evidence of the impact of CPOE design aspects on clinicians' ordering behavior and opportunities for error attributable to the interaction process. We reviewed the literature for original studies describing a (usability) evaluation of CPOE systems' design aspects and report on the preliminary results. Recommendations are provided to enable CPOE system designers to create systems that are not only more user friendly and efficient but also safer.

1. Method

A systematic literature search in PubMed, EMBASE and Ovid MEDLINE of papers published from 1986-2006 was performed. These databases were searched for English-language publications with four groups of key terms related to: (1) CPOE and Electronic prescribing systems, (2) Computer and electronic, (3) Prescription and, (4) Usability and workflow. Key terms in each group were combined by operator "OR". Generally we used the following combinations in search strategies; (1 AND 4) OR (2 AND 3 AND 4) to extract relevant studies.

The authors reviewed the abstracts of the resulting papers independently. Articles were included when they described an original usability evaluation study of a CPOE medication system in health care. In case of any doubt, full texts were reviewed. Editorials, letters, and conceptual papers were excluded. Any disagreements between researchers were resolved through discussion. We clustered CPOE design aspects described in these papers into six groups as follows: 1) timing of alerts, 2) log in/out procedures, 3) pick lists and drop-down menus, 4) clues and guidelines, 5) documentation and data entry options, and 6) screen display and layout. Selection of design aspects and of their effects on CPOE systems' usability and clinicians' ordering patterns was done by both authors.

2. Results

Our search in PubMed and Ovid resulted in 724 papers published from 1986 to 2006 of which 11 papers met our inclusion criteria and were reviewed for data collection (see Table 1). Published studies [5;8;9] reported that alerts which show up too early or too late in the workflow of CPOE users ordering medication can lead to errors from which users cannot recover. Users may indeed search for alert information at a different moment in time unnecessarily hampering and prolonging the ordering process [8].

Post-hoc alerts persuaded users to shift the responsibility of drug interaction checking to the pharmacist [5]. In another study [10], failure to alert in the proper time caused deactivation of orders and prevented users to be aware of a drug-allergy interaction. Lack of timely duplicate checking when clinicians ordered a new dose of the same medication, [11] or the same medication in another form, or re-ordered medication prescribed earlier [4] resulted in duplicate medication orders, potentially leading to overdoses. Likewise, a failure to warn CPOE users that antibiotic drugs had to be preapproved caused delays in approval, and resulted in gaps in antibiotic therapy [5].

Inconvenient logging procedures, especially when the log-out takes time because of security measures, incited many physicians to order medications at computer terminals not yet “logged out” by other physicians [5;12]. As a result, physicians signed orders that they did not enter themselves. Using another clinician’s logged-in session can yet result in either unintended patients receiving certain medications or patients not receiving the intended medication [5].

Picking wrong items from drop down lists and multiple choice items on the computer screen, and failure to differentiate look-alike patient names led to selection errors [7;10;11], e.g. wrong patient, a wrong drug, and wrong drug routes. Close proximity of selection items on the screen e.g. items on the drop down list for order routes may cause juxtaposition errors [9]. Lengthy lists of items in menus, with few of

Table 1. CPOE design aspects in the relevant articles

Reference number	Alerting			Log –in/ out	Drop-down lists	Clues and guidelines					Documentation and data entry components			Screen display/ layout and appearance						
	Wrong timing of alerts	Real time alerts	Failed alerts	Lack of Alert functionality	Inconvenient log-in procedures	Lengthy list of menu items	Proximate screen items	Screen icons	Lack of clues and information	computation	Obscure order hierarchies	Poorly designed icons	Documentation templates	Data entry fields	Pre-set global schedules	Case sensitive fields	Poorly conceptualized graphical representations	Multiple screen displays of medications	Faulty screen display	Suboptimal displayed measures
3									✓											
4				✓										✓					✓	
5	✓			✓	✓													✓	✓	✓
7						✓					✓						✓		✓	
8	✓	✓							✓					✓			✓			
9	✓						✓													
10			✓			✓									✓					✓
11				✓		✓													✓	
12					✓								✓							
13						✓		✓						✓					✓	
14								✓			✓	✓								
Frequency	3	1	1	3	2	4	1	1	2	1	1	1	2	2	1	1	2	1	5	2
Total	8			2	5	6					6			10						

the items visible at once, are difficult to use [13] and require users to scroll down to see the other items [7;13].

Icons on the screen reminding users of forthcoming tasks help them to organize and time their tasks. Nurses likewise can be confused by CPOE systems displaying exact times for drug administration without providing clues as to whether these specific times are critical or not. Horsky et.al [3] found that screens providing few clues and insufficient information to support users in their natural task flow necessitates users to perform a series of demanding estimations and comparisons required to accomplish the order tasks. Although automated dose calculation facilities can assist users in deciding on a drug dose, computations that are represented without their algorithmic basis forces users to calculation which complicates the interaction [8]. Obscure hierarchical structuring of orders require novice CPOE users to involve in a prolonged trial and error task causing time delay, selection of wrong drug sets and failures to find an appropriate drug order set [7].

Banet et.al [14] reported that documentation templates prompting users to enter certain information granted efficiency and standardization of documentation e.g., use of these templates prevented double/triple charting. CPOE interaction structure relied upon a cognitive model of classifying orders which the physicians did not always share, introducing difficulty in structured data entry and prolonging this procedure[12]. The meaning of a number of fields specially adjacent ones in a data entry screen can be easily misinterpreted [4;8]. These kinds of misinterpretations can, in a positive sense generate alerts prolonging the ordering process. One study [13] showed that the use of grey boxes for highlighting preferred time-slots for drug dispensing by nurses that were to be activated by physicians were misinterpreted by the same physicians as fields in which no data could be entered. Another study [10] showed that string sensitivity of data fields ('TID' entered instead of 'T.I.D.') during ordering caused ordering failure.

Several studies [4;5;7;8;10;11;13] reported on the effects of suboptimal screen displays of medication ordering systems. Poorly conceptualized graphical representations make it difficult for CPOE users to find certain information leading to inefficient searches for information provided by the system. Poor conceptual presentation of alerts likely increase cognitive effort and require users to engage in an extensive search for this information, unnecessarily prolonging the ordering process with potential for medication errors[8]. Moreover poor display of entered orders do not allow for simple visual reviews of these orders [8], necessitating users to scroll through several screens and forcing them to rely on the history of orders in their own memory [7], which creates opportunities for medication errors. Multiple-screen displays of a patient's medications indeed prevented physicians from seeing a patient's complete medication record, with medication discontinuation and selections of wrong medications as result [5]. Huge amounts of information displayed in one screen likewise forces clinicians into a cognitively exhausting and error prone task[7]. Fragmented CPOE displays for example made it difficult for physicians to identify the patient they were actually ordering for [5;7]. Subtle differences in lay-out and appearance of screen entry forms, data labels and values yet providing important functional differences led to erroneous interpretation of stop times for drugs [4]. Moreover, lack of an explicit indication that a laboratory result was not from the same day as the day of the drug order let a clinician set out another drug order, leading to an overdose [4]. While another study found that invisibility of dates on printouts confused users who assumed that two printouts for a similar medication represented a duplicate order whereas in reality these concerned two different orders[11]. Suboptimal labeling (for example for 'package' instead of 'tablets') of medication amount [10] and displaying drug dosages according to pharmacy warehousing and purchasing decisions

rather than according to clinical guidelines [5] can lead to overdoses or medication under-dosing.

3. Discussion

The outcomes of this review shed light on design features of CPOE medication systems and concealed usability problems related to them. Owing to the characteristics of CPOE design aspects described in this paper, designers can consider them from the start point to avoid suboptimal CPOE usability. Our preliminary results show that in designing CPOE systems, the following points should be considered to avoid error-prone CPOE designs: 1- Design interfaces that explicitly map to the workflow patterns of clinicians, so as to keep the ordering process as less cognitively complex as possible. 2- Provide users by clues in the interface to optimally support them in medication ordering. These external cues in the display can fulfill a central role in controlling the CPOE user interaction. 3- Reduce the layers of screens (to a maximum of 3 layers) to facilitate users navigation in the system. This recommendation is based on insights from cognitive psychology[15], and consistent with the study results showing that for certain CPOE systems clinicians report a loss of overview when they are forced to navigate through too many different screens to review a patient's current medication status. 4- Alerts should be timed properly in order to be displayed in a timely manner; that is at the moment a clinician would himself search for this information. 5- Alerts should also be displayed in a more prominent position on the screen, so that these are more easily noticeable. 6- Use consistent terms throughout the CPOE system that are related to task. 7- Group screen elements that are related, physically together, so that the layout of these elements on the screen guide the CPOE user to the information they are looking for. 8- Organize screen-elements into logical groups, visually separated by space and alignment. 9- Give enough space to user interface items to prevent users from inadvertently clicking the wrong options. 10- Make active and passive screen elements easily distinguishable by consistent use of tick boxes and pick lists. 11- Use colors sparingly and consistently, giving important elements (e.g. alerts) prominence through contrast, making it easier for clinicians to notice information intended to arrest their attention.

In the same way, results showed that in order to have CPOE systems aligned with workflow pattern of clinicians the following designs should be avoided: 1- Deep navigational structures in order to make the user system interaction more effective. 2- Too close positioning of Screens elements. 3- Use of the same color for data entry fields and fields in which no data can be entered. 4- Long drop-down menus which require the user to scroll down. 5- Use of documentation templates classifying orders different from the cognitive model of physician ordering. 6- Use of string sensitive data fields especially for abbreviations. 7- Provision of too many alerts. To prevent alert overriding prioritize important alerts by colors. 8- The displaying of huge amount of information in one screen to avoid an exhausting and error prone task by physicians. 9- Obscure hierarchies of orders and order sets.

Since a CPOE user interface handles the communication with the end-user, thoughtful considering its design is crucial to produce a usable system.

4. Conclusion

To our knowledge up to now, this is the first review focusing exclusively on design aspects of CPOE medication systems and their potential for introducing error in the order process. Despite of our extensive search we ended up with few original usability studies about CPOE medication ordering systems indicating that more research should be done in this respect. Characterization of consequences associated with certain CPOE design aspects provides insight into how CPOE system designs can be improved to enhance physicians' adoption of these systems and their success. We ourselves will at least refer to these insights as input to a usability evaluation of the CPOE medication system of the Academic Medical Center (AMC) in the Netherlands with the ultimate aim to improve its design so as to prevent medication errors.

Reference List

- [1] D.S.Bell, S.Cretin, R.S.Mariken, and A.B.Landman, A conceptual framework for evaluating outpatient electronic prescribing systems based on their functional capabilities, *J.Am.Med.Inform.Assoc.* 11 (2004) 60-70.
- [2] J.M.Teich, J.A.Osheroff, E.A.Pifer, D.F.Sittig, and R.A.Jenders, Clinical decision support in electronic prescribing: recommendations and an action plan: report of the joint clinical decision support workgroup, *J.Am.Med.Inform.Assoc.* 12 (2005) 365-376.
- [3] J.Horsky, D.R.Kaufman, and V.L.Patel, When you come to a fork in the road, take it: strategy selection in order entry, *AMIA.Annu.Symp.Proc.* (2005) 350-354.
- [4] J.Horsky, G.J.Kuperman, and V.L.Patel, Comprehensive analysis of a medication dosing error related to CPOE, *J.Am.Med.Inform.Assoc.* 12 (2005) 377-382.
- [5] R.Koppel, J.P.Metlay, A.Cohen, B.Abaluck, A.R.Localio, S.E.Kimmel, and B.L.Strom, Role of computerized physician order entry systems in facilitating medication errors, *JAMA* 293 (2005) 1197-1203.
- [6] L.W.Peute and M.W.Jaspers, The significance of a usability evaluation of an emerging laboratory order entry system, *Int.J.Med.Inform.* 76 (2007) 157-168.
- [7] J.Horsky, D.R.Kaufman, M.I.Oppenheim, and V.L.Patel, A framework for analyzing the cognitive complexity of computer-assisted clinical ordering, *J.Biomed.Inform.* 36 (2003) 4-22.
- [8] J.Horsky, D.R.Kaufman, and V.L.Patel, Computer-based drug ordering: evaluation of interaction with a decision-support system, *Medinfo.* 11 (2004) 1063-1067.
- [9] J.S.Ash, D.F.Sittig, R.H.Dykstra, K.Guappone, J.D.Carpenter, and V.Seshadri, Categorizing the unintended sociotechnical consequences of computerized provider order entry, *Int.J.Med.Inform.* 76 Suppl 1 (2007) 21-27.
- [10] C.Zhan, R.W.Hicks, C.M.Blanchette, M.A.Keyes, and D.D.Cousins, Potential benefits and problems with computerized prescriber order entry: analysis of a voluntary medication error-reporting database, *Am.J.Health Syst.Pharm.* 63 (2006) 353-358.
- [11] V.M.Bradley, C.L.Steltenkamp, and K.B.Hite, Evaluation of reported medication errors before and after implementation of computerized practitioner order entry, *J.Healthc.Inf.Manag.* 20 (2006) 46-53.
- [12] C.H.Cheng, M.K.Goldstein, E.Geller, and R.E.Levitt, The Effects of CPOE on ICU workflow: an observational study, *AMIA.Annu.Symp.Proc.* (2003) 150-154.
- [13] M.C.Beuscart-Zephir, S.Pelayo, F.Anceaux, J.J.Meaux, M.Degroisse, and P.Degoulet, Impact of CPOE on doctor-nurse cooperation for the medication ordering and administration process, *Int.J.Med.Inform.* 74 (2005) 629-641.
- [14] G.A.Banet, D.B.Jeffe, J.A.Williams, and P.V.Asaro, Effects of implementing computerized practitioner order entry and nursing documentation on nursing workflow in an emergency department, *J.Healthc.Inf.Manag.* 20 (2006) 45-54.
- [15] Gardiner, M. M. and Christie, B. Applying cognitive psychology to user interface design, John Wiley & Sons, Chichester, 1987.

Decision Support System Supporting Clinical Reasoning Process – an Evaluation Study in Dementia Care

Helena LINDGREN¹

Department of Computing Science, Umeå University, Sweden

Abstract. In this paper, a case study is presented in which an early prototype of a decision-support system was integrated in the process of investigating patients with suspected dementia and evaluated. The aims were to capture and model the complex target activity for the purpose of knowledge acquisition and formalization, and qualitatively evaluate the system's compliance with reasoning and work processes as part of the development of a decision-support system for the domain. The results show that contextual factors such as local routines in clinical practice motivate further development of the support integrated in the system for establishing preliminary diagnoses in the investigation process.

Keywords. Activity theory, Work analysis, Knowledge-based systems, Decision support, Assessment-Evaluation, Process, Compliance, Human Factors, Cognitive tasks, User-computer interface, Knowledge management, Dementia care

Introduction

The clinical decision-support system (CDSS) DMSS (Dementia Management Support System) is being developed for assisting medical personnel in the investigation of suspected cases of dementia [1]. The application domain is used to investigate a range of issues concerning knowledge acquisition, knowledge formalization and representation, interaction design and methods for transforming informal clinical practice knowledge into usable support in a CDSS. The main purpose of the system is to function as an extension of the individual actor's cognitive ability and as a common ground for collaborative and distributed team work. The system is designed to support higher-level cognitive functions such as reasoning, decision-making and learning. The support is based on established and international clinical guidelines and classifications.

A model of the main reasoning process is integrated in the system, which is based on how experts report that they proceed in the investigation of dementia. The model is validated against clinical guidelines and qualitatively with physicians in different clinical domains that are experts on dementia [1]. However, previous evaluations have

¹ Corresponding author: Helena Lindgren, Department of Computing Science, Umeå University, SE-901 87 Umeå, Sweden; E-mail: helena@cs.umu.se.

been limited to snap-shots of current patient cases. For the purpose of investigating the prototype system's compliance with actual investigation processes, a case study was conducted in a clinical setting. The results of the study mainly consist of two parts; 1) a general model of the activity in focus (investigation of dementia) as conducted in this particular clinic, and by the particular expert physician participating in the study; and 2) the formative evaluation of the system related to the activity analysis. This paper presents the second part, which embraces reasons for breakdown situations, and knowledge representation and interaction design issues to be handled before further evaluations are conducted.

1. Methods, Materials and Procedure

An action research approach is taken throughout the DMSS project with the cultural-historical activity theory as theoretical framework for analysis and development [2, 3]. Among other tasks, the knowledge acquisition process contains the central activity of collaborative knowledge building (CKB). In the project, a method similar to [4] is applied. Reflective conceptual artifacts, sometimes integrated in prototypes, are used to create a common understanding of knowledge components to be integrated in the system. Whenever the knowledge mediated through the design of the user interface does not comply with the shared understanding, or the use of the system highlights misunderstandings, the iterative process of CKB continues. These situations are commonly denoted as *breakdown*, or *conflict* situations in terms of activity theory, which are carefully analyzed and the results are used for improving the system.

In our case study, an early version of the prototype system DMSS was integrated as a prompting tool for discussions about content and design of the system. However, the main focus for the case study and unit for analysis was the investigation process as a whole.

Five patient cases under investigation for suspected dementia were the focus for this work. The studies were conducted in the setting of an out-patient ward at a geriatric clinic which investigates suspected patient cases with cognitive diseases. The patients were selected by including all new cases who were willing to participate in the study and who were assigned a particular geriatrician from a certain date. The geriatrician is considered an expert in dementia care with more than 15 years of experience, and had participated in CKB activities before the case study. Consequently, the geriatrician was expected to be familiar with the purpose of the DMSS project and the content of the system.

We use the term *investigation* to denote the sum of diagnostic actions, interventions and follow-up actions, executed at this particular clinic. The time period of investigation for each patient was in the range of 3-10 months.

The investigation process was documented in the electronic health record, and by observations of clinical encounters and interviews. An approach inspired by grounded theory was used as the method for analysis in order to identify general purposes and structures of activity, however, this was combined with the theoretical framework of activity theory. A model of the clinical activity was formed, which was used as the framework for interpretation in the analysis of breakdown situations that occurred in the use of the system.

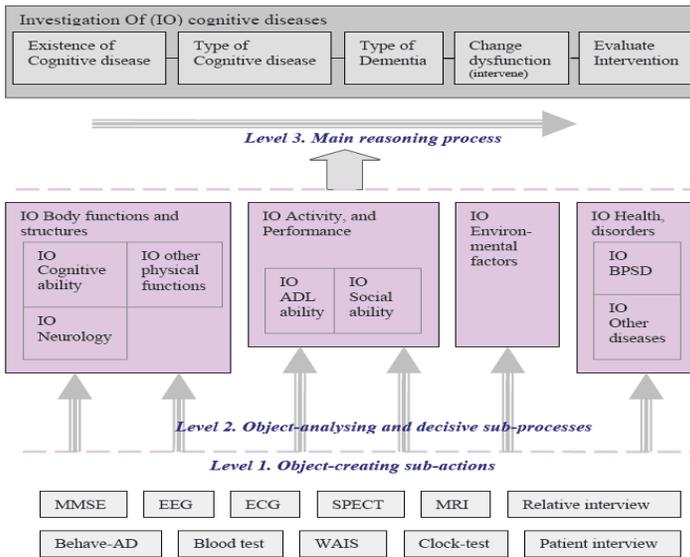


Figure 1: Different levels of complexity in the process of investigating cognitive diseases. BPSD is an abbreviation for Behavioral and psychological symptoms in dementia, and MMSE, EEG, ECG, WAIS, SPECT, MRI and Behave-AD are abbreviations for examinations.

2. Results

The model of the clinical investigation activity consists of three parts; an activity system model, a patient model and a process model. The *activity system model* is based on the work of Engeström [2] and includes resources (e.g., knowledge, guidelines, tools) and constraints (e.g., routines, work division) built in the organisation and in individuals, affecting the way an activity is executed. The *patient model* is a semi-formal conceptual model of the object of investigation, i.e., the patient’s health and well-being, based on clinical terminologies and classifications useful for the domain.

The *process model* of the clinical activity includes typing of actions based on their purposes, their relations and three main levels of complexity, which also correlate to three analytic perspectives on the activity. In the analysis of actions with a focus on their different purposes from a physician’s perspective, three parallel processes were identified of which two constitute reasoning procedures including medical decisions. We distinguish between the two reasoning processes directed by different purposes (which correspond to Level 2 and 3 in Figure 1); 1) the physician’s examination of specific domains and 2) the main reasoning process. Simplified, the first purpose is to value functionality per se and the second purpose is to value the reasons for function or dysfunction in the perspective of dementia diagnosis and intervention. Further, a distinction is made of the logistic, or administrative process, which is imposed by the rules governing the whole process. These three views of the process (logistic process, investigation of specific domains and the main reasoning process) are executed in parallel, they are cross-fertilizing and dependent on each other, and are partly overlapping depending on circumstances in the environment. In our case study, conflicts between these perspectives were identified as causes of breakdowns in the use

of DMSS. Other causes of breakdowns were generated by conflicts within the activity system and within the user.

The decision-support system DMSS is primarily designed to support the main reasoning process towards establishing diagnosis and guide the process of investigating specific domains (Figure 1). The main reasoning process integrated in the system is based on guidelines and on domain experts' descriptions of how they execute their reasoning. Investigation of certain core (necessary) features is required by the system at different stages in the process, while the investigation of supportive features is suggested when appropriate. The initial status of all features are unknown by default.

In the case studies, the system was used immediately after the patient's first encounter with the physician. In four of the five cases, the system did not guide further than the first step at this point, in which the existence of cognitive disease is assessed. The reason was that the system required, based on certain guidelines, that the possibility of exposure to toxic substances should be investigated and whether the cognitive ability was fluctuating during the course of a day, before proceeding in the reasoning process to differentiating between diseases. In the fifth case the physician concluded that the patient did not have fluctuating cognitive ability. This was based on that nothing in what the patient or the patient's next-of-kin told suggested the contrary. The reason why the particular physician who participated in the study did not explicitly investigate these factors during the first encounter, was explained to be that the time scheduled for the encounter is limited and because the pressure on the patient and the next-of-kin has to be limited to what is the most important in the first visit.

However, the analysis also reveals that although evidence is missing that these factors are not manifested in the patient, the physician in some cases concluded their absence based on that neither the patient nor the relative provided any suggestions that these features were present. In practice, the physician draws conclusions beyond the constraints integrated in the DMSS system and beyond the current interpretation of the clinical guidelines in the first encounter by formulating a preliminary diagnosis based on the limited evidence at hand.

In the case when the physician proceeded with the use of the system, the system suggested a typical vascular dementia, based on a certain clinical guideline. The physician however, assessed Alzheimer's disease with vascular symptoms, which is a diagnosis in accordance with another clinical guideline. This imposes a new analysis of how the valuation should be done of Alzheimer's disease vs. other dementia diseases in typical cases.

The system reveals what information is missing, in order to guide the physician in further investigations before the reasoning process can be continued. In the case studies, the physician acknowledged and agreed with the response from the system, that no reliable diagnosis could be assessed because of the reasons the system presented. Nevertheless, the physician assessed a preliminary diagnosis, which was documented in the electronic health record. This was done without using the overview provided by the system of the evidence at hand, which could have been useful. By the time of the second encounter with the patient, further investigations had been made and more evidence was gathered from a neuropsychologist's examinations and radiology investigations. At this point, the expert physician's need for a decision-support system was minimal, since the evidence at hand was sufficient to draw conclusions of diagnosis and establish a plan for interventions in these particular patient cases. The additional investigations were used to confirm or reject the preliminary diagnosis assessed at the first encounter in the case studies.

3. Discussion and Conclusions

The study presented in this paper is limited, in that only one physician participated in the study and a limited amount of patients. However, what can be concluded based on the case study is that the support provided by DMSS at the first encounter with the patient is limited and unsatisfactory, primarily because of the constraints of the logistics in the process. The establishment of a preliminary diagnosis was done without the supplementary support from the system that could have been provided. The reason for this was judged to be the system's heavy emphasis on the missing information that prevented the reasoning process to proceed towards a final diagnosis according to guidelines, which shadowed the value of existing evidence and knowledge related to hypothesis building. Further, to what extent the system could provide support in these cases by changing the formal representation of the critical factors identified in the study to be absent by default, if no signs prompts the physician to investigate their presence, needs to be analysed in the perspective of domain knowledge. Therefore, further development of the different levels of reliability of diagnostic decisions is needed, i.e., for hypothesis building and final diagnostics.

Another observation was that the expert physician had minimal use of the system at the second encounter with the patient in these particular cases, since the final diagnosis was established beforehand and the physician perceived all but one case as typical, uncomplicated cases. However, in other evaluations, expert physicians found the system highly useful for verifying their diagnoses in difficult cases [1]. Furthermore, physicians with minor experience found the system useful as a checklist and guide also in "easy" cases. Therefore, more patient cases and physicians with different expertise and amount of experience need to be included in future evaluations of the system's compliance with work processes in order to obtain a general view of the benefits of the system. Additionally, the potential the system has as a tool for developing expertise in clinical practice will be investigated.

Another reason why the system did not fully comply with the physician's reasoning process, could be that the reasoning executed by the physician is probably more directed by what particular evidence is gathered at a certain time point, which may prompt the physician to follow lines of investigation, a behaviour commonly seen in experts [5]. Whether the particular lines seen in this study are limited to this particular expert physician remains to be investigated.

To summarise, reasons for the system's unsatisfactory compliance with the investigation process actually taking place can be many-fold. One is the logistic routines with limited time for patient encounters and waiting time for radiology examinations. Second, the domain knowledge is ambiguous and can be interpreted in different ways, which leaves the physician to choose how to proceed on a case-by-case basis. The CDSS needs to be flexible and adjustable in providing support according to local and individual preferences of clinical guidelines, etc. Third, observed clinical evidence is not valued in a consistent way in all cases and in all situations by the individual expert physician. This is due to interrelationships between evidence, which affect the importance of each piece of evidence, depending on different contexts of interpretation. This becomes more clear by the distinction of the two-dimensionality of the investigation process where one dimension can be seen as the gathering of data and interpreting the data into information, which is a process done with the general perspective of (healthy) function vs. dysfunction (Level 2 in Figure 1). This process is orthogonal to the process of diagnostic reasoning and intervention, to which the

information are fed as evidence in the reasoning towards diagnosis and appropriate choice of intervention (Level 3 in Figure 1). Data is interpreted into evidence differently depending on which context is the primary context for reasoning, and a CDSS needs to support this distinction with appropriate formal representations and a well-founded design of the interactive clinical reasoning, including support for hypothesis generation.

Finally, some of the tacit knowledge that develops with experience can be expressed in sessions of knowledge building. However, some is revealed only in observations of daily work, as is evident in this case study. Therefore, qualitative studies of CDSS in clinical use as part of the development process are essential in order to develop decision-support systems that have the potential of becoming useful and used in daily practice.

4. Summary and Future Work

A case study in a clinical practice setting was conducted in order to investigate and evaluate a decision-support system's compliance with the main reasoning and investigation process of a suspected dementia case. The study design was qualitative and formative, in order to obtain results that can inform the development process and be reused in future evaluations as a baseline frame of reference. The model of clinical activity that is developed is used as a dynamic conceptual tool for assessing local clinical routines and practices in the development process. The model is also used as a frame of reference in the presented evaluation of the prototype system DMSS.

Based on the case study, the development of DMSS continues with the integration of support for hypothesis generation, support for handling complex patient cases with rare diagnoses and/or additional behavioural and psychological symptoms, and extended evaluations of the system integrated and used in clinical practice.

Acknowledgements

We thank patients and physicians who have participated in evaluations. The DMSS-project is partly funded by Swedish Brain Power.

References

- [1] H. Lindgren, *Decision Support in Dementia Care – Developing Systems for Interactive Reasoning*, <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-1138>, Umeå University, 2007.
- [2] Y. Engeström, *Learning by expanding: An activity-theoretical approach to developmental research*. Helsinki: Orienta-Konsultit, 1987.
- [3] L. Vygotsky, *Mind in society: The development of higher psychological processes*, Cambridge: Harvard University Press, 1978.
- [4] G. Singh, Investigating the support of reflective activities by collaborative technologies: an activity theory based research model, In *Proc. TT211C2006* (2006), 49-58.
- [5] V. Patel, D. Kaufman, J. Arocha, Emerging paradigms of cognition in medical decision-making, *Journal of Biomedical Informatics* **35** (2002), 52-75.

Evaluating Inter-Professional Work Support by a Computerized Physician Order Entry (CPOE) System

Zahra NIAZKHANI^{1a}, Habibollah PIRNEJAD^a, Antoinette de BONT^a, Jos AARTS^a
^aInstitute of Health Policy and Management (iBMG), Erasmus University Medical Center, Rotterdam, the Netherlands

Abstract. Physician-centered design for computerized physician order entry (CPOE) systems overlooks the collaborative, multi-professional nature of medical work. We analyzed the compatibility of the conceptual model of inter-professional workflow underlying a CPOE system with real-life workflow in the medication ordering and administration process. We conducted twenty-three semi-structured interviews with key informant users and analyzed the handwritten documents and computerized printouts used in daily work in a Dutch academic medical center. The interview transcripts were analyzed on the basis of three conceptual themes in the inter-professional workflow: division of tasks, flow of information, and task coordination. The CPOE system fundamentally reorganized the existing work procedures of the three professional groups involved, mainly by reassigning tasks and by reallocating areas of expertise. Although the system improved the flow of medication-related information from physicians to nurses or pharmacists, this flow was only in one direction; the system did not allow information transactions in the reverse direction. It also failed to coordinate the medication-related tasks of professionals from different disciplines. To maintain the necessary level of coordination, the professionals had been obliged to consider additional methods of communication, such as phone calls or face-to-face discussion. We identified several workflow integration issues after the implementation of a CPOE system. Our insights into these issues can help ensure that system design or redesign properly integrates all professional groups' tasks, information, and areas of expertise into those of the physicians. Only then can these systems support the actual inter-professional workflow in the medication process.

Keywords. Computerized physician order entry, workflow, collaboration

Introduction

Implementing computerized physician order entry (CPOE) systems has proved to be difficult [1]. This is mainly because the deployment of these systems may disrupt medical workflow, and possibly trigger resistance in users [2]. It has been shown that the workflow model embedded in these systems does not match actual workflow between professionals [3]. Gorman et al. criticized these systems for their linear,

¹ Corresponding author: Zahra Niazkhani, MD, MHIM; Erasmus University Medical Center, P.O. Box 1738, 3000 DR, Rotterdam, The Netherlands; e-mail: z.niazkhani@erasmusmc.nl

sequential, and unidirectional model of care processes [4]. It has also been pointed out that a narrow focus only on physicians may overlook the collaborative, multi-professional nature of medical workflow, which involves not only physicians but also professionals such as nurses and pharmacists [5]. Despite these shortcomings, the literature to date has nonetheless paid little attention to how well the design of these systems takes account of the multi-professional and interdependent nature of medical workflow. Neither is it completely clear which aspects of inter-professional workflow are most affected by CPOE systems.

To analyze the compatibility of the workflow model underlying this system with real-life inter-professional workflow, we conducted a qualitative study of a computerized medication order entry system implemented in a Dutch hospital. For this purpose, we examined the role of the system in integrating the work of one professional group with that of others in the medication process. In particular, we were interested in identifying areas in inter-professional medication-related workflow which are either supported by the CPOE system or rendered difficult by it. The resulting insights can help designers and implementers to consider redesigning both systems and care processes, thereby creating a better fit between the system and the multi-professional nature of the medication process.

1. Conceptual Framework

Our study was inspired by Wears and Berg, who pointed out that “many of the difficulties do not result from bad parts of the systems but are inherent in the perspectives and theories of medical work (and the role of IT in this work)” on which these systems are founded [6]. Drawing upon studies of medical work carried out in the social sciences and in the field of computer-supported cooperative work, we identified three conceptual themes that were relevant to the workflow between professional groups using information systems: division of tasks, flow of information, and task coordination [7-9]. The medication ordering and administration process involves various professional groups whose tasks are extremely interdependent. Avoiding possible conflicts requires an effective division of labor that not only takes account of different professional groups’ work domains, but also coordinates them effectively [7].

While this calls for an accurate and timely flow of information between the groups [8], it is not sufficient, as the *ad hoc* nature of the medication process can cause medication plan to change frequently. Various professional groups therefore are required to coordinate their interrelated tasks and to ensure a shared understanding of the medication plan [9]. Figure 1 shows the relationship between these three concepts in our conceptual model.

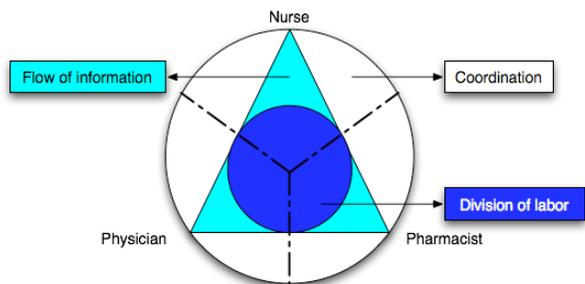


Figure 1. A conceptual model for inter-professional workflow in the medication process

2. Methods and Materials

In 2001, a commercially available computerized medication order entry system (Medicatie/EVS) was implemented at Erasmus University Medical Center, an academic hospital in the Netherlands. Medicatie/EVS was chosen for implementation because it was compatible with the hospital's existing information systems. When physicians use this system for order entry, nurses receive printed medication-order-labels, which they stick into a paper-based, patient medication administration record (MAR). Nurses use the system to request non-stock drugs from the pharmacy.

The first and the second authors conducted 23 semi-structured interviews with a sample of key informant users (i.e., 12 nurses, eight physicians, two pharmacists, and one pharmacy technician) between November 2006 and April 2007. The interviews were performed in the interviewee's working environment where they could show us how they work with the system. The nurses and physicians interviewed were recruited from both internal and surgical specialties. During the interviews, we focused on the role of the system in integrating one group's work with that of others. The interview transcripts were analyzed on the basis of the three themes discussed above. In addition, to examine their role in the workflow, we analysed handwritten documents (e.g., MAR), and system printouts (such as the patient medication overviews) used in daily work.

3. Results

The results are presented on the basis of the three themes. Per theme, we particularly focused on workflow between two or more professional groups.

3.1. Division of tasks

The CPOE system in our study fundamentally reorganized the existing work procedures, affecting the workflow between the three professional groups both advantageously and disadvantageously.

By forcing strict levels of authorization for executing tasks, the system reinforced professional boundaries. In some cases, this was perceived as beneficial. Nurses, for example, were no longer questioned about the changes in orders, as, in the new situation, changing medication orders were not considered part of a nurse's responsibility. In other cases, however, this reorganization negatively affected workflow by reallocating areas of expertise and by reassigning tasks.

Although it removed order decryption and transcription tasks for nurses, the concept of physician order entry in this system enforced a central position for physicians. This meant that physicians were sometimes forced to decide on the details of orders beyond their areas of expertise. One physician, although was generally satisfied with the system, noted: "When you have to put 10 prescriptions, then you have to check for all [details], [for example] let's go to IV: IV wide, IV peripheral, IV central... or just IV; it doesn't matter... these are very specialized." (Feb. 2007)

In turn, this centralized decision-making violated nurses' work domain: nurses experienced difficulties in implementing physicians' detailed medication orders, particularly with regard to time and route of administration. For example, they often had to adjust the time of drug administration to fit in nursing work routines. Because

they had not been authorized to change medication orders using the system, nurses manually registered these adjustments on the order labels. This worked well only for available drugs in the ward stocks. However, problems arose when such adjustments required nurses to request drugs from the pharmacy. For instance, before the implementation of the CPOE, nurses could, on their own initiative, change a patient's IV antibiotics to oral forms after three days of infusions. After the implementation, however, they had to wait for physicians to change the order in the system, because the pharmacy would reject the nurses' drug requests in the absence of the electronic orders.

3.2. Flow of information

This system improved the flow of medication-related information from physicians to nurses and to the pharmacy. This was especially the case due to legible and complete, single medication orders and could save many call-backs to physicians. However, nurses received piles of order labels after physicians entered their orders into the system. This required nurses to sort the labels out per patient and put them into the right administration-records.

Sometimes, the uniform black and white structure of these labels caused nurses to mix them up for different patients. To ensure an accurate flow of information, nurses therefore had to ensure that they had the right medication labels for the right patients. For this purpose, the nightshift nurses were obliged to double-check each patient's order labels with a medication overview printed out from the system.

Unfortunately, the improvement in information flow was only in one direction; the system did not allow information transactions in the reverse direction. Because nurses recorded medication-related information in the paper-based nursing records, the medication data became fragmented in the electronic and the paper-based systems. Similarly, because nurses and pharmacists were not allowed to input into the system, information flow from other professional groups to physicians through the system was inadequate. For instance, physicians had practically no easy access to the MAR, which was available in the nurses' working station. To develop comprehensive, integrated patient medication information, this information therefore had to be communicated directly between physicians and nurses during the medical rounds.

Finally, the flow of information between nurses and the pharmacy was insufficient. To acquire the information they needed, both groups had to call each other.

3.3. Task coordination

The medication-related tasks among professionals were coordinated by other methods of communication and not only through the system. None of the professional groups actually counted on the system for this purpose. While phone calls played an important role in coordinating interdependent tasks between professionals from different services (such as physicians and pharmacists, or nurses and pharmacy technicians), physicians and nurses who worked closely together depended most on face-to-face communication.

In all the specialties interviewed, physicians and nurses discussed the overall medication plans in medical rounds, during which the majority of the decisions on changing medication plan were made. For reference, nurses often made notes on these decisions. Without these rounds, no shared understanding of the medication plan could be developed; both groups therefore depended on the discussions in these rounds.

If a change was necessary in the evening or night shifts, physicians would have to inform nurses directly. Relying merely on the system and on the printed labels to coordinate these changes was perceived as risky. In fact, if a medication order was lost among other papers, or if there was a printer failure, none of the nurses or physicians would realize on time. On the other hand, as one senior head nurse noted, when nurses received any new order label, they often contacted the prescribing physician: "...in such a case [a change in medication plan], physicians usually tell us; otherwise, if we see there is a controversy between the medication-label and our notes, then we [will] call physicians and ask for the reason". (Jan. 2007)

Similarly, the procurement of non-stock drugs required nurses to take extra coordinative steps beyond the system. Because the system was not available at the bedside, physicians entered the orders in their offices later on. Due mainly to the time pressure caused by other clinical duties after their medical rounds (such as operations or outpatient visits), they often delayed entering their orders. As a result, nurses were able to send the electronic requests to the pharmacy only later during the day.

To notify the pharmacy technicians of new requests sent during the afternoon, nurses also had to call them personally, because, due to internal policy at the pharmacy, the technicians checked the electronic requests only twice a day: at 8 and 12 o'clock. In order to emphasize the necessity of drug delivery for the same day, this coordination redundancy –referred to almost by all the nurses interviewed– had become routine.

4. Discussion and conclusion

Our study revealed several workflow integration issues. Besides advantages such as legible and complete medication orders, which have resulted in overall satisfaction with the system implemented, we identified instances in which the system has inappropriately integrated professionals together by reassigning tasks and reallocating the areas of expertise. This system also caused patient medication-related information to become fragmented in both the paper records and in the electronic records, and also in different professional domains. Neither did the system support professional groups in the temporal coordination of tasks, nor in making sense of new changes in the medication plan. To integrate work, they frequently bypassed the system or added new steps (e.g., double-checking the orders) and extra coordinative tasks (e.g., phone calls).

Our study shows that the system challenged the effective integration of various professional groups' work not only by reorganizing the areas of expertise, but also by reinforcing strict boundaries around professional domains. In fact, the workflow model underlying this system overlooked the overlaps that normally exist in practice. These findings are similar to those of other studies which have emphasized that CPOE systems may limit opportunities for decentralized decision-making in the medication process [10]. In our study, to compensate for such limitations, the professionals tried to bypass the system and to adjust the orders on the basis of their own work organization, although it was not easy to provide other parties with feedback on these adjustments. Even in a highly advanced CPOE system, a similar lack of effective integration has been found between nursing processes and the computerized system [11].

Negotiation between co-working professionals is critical to creating a shared sense of a care plan and to adjusting the work of one professional group with that of another. In our study case, these purposes were served by the medical rounds, which enabled physicians and nurses to negotiate their overall medication plans. Nevertheless, this

was unhelpful with regard to the details of orders and also to the changes which were made beyond rounds: as stated above, extra communication methods such as phone calls and face-to-face communication had to supplement the system. Similar coordination redundancies have been reported in another CPOE study [3].

With regard to inter-professional medication workflow, implementing a CPOE system is a double-edged sword. Our study not only contributes to a deeper understanding of the interdependent nature of medication-related tasks among professional groups working in the same or different services, it also identifies where the problems lie with the CPOE system implemented. Our findings are in accordance with the argument of Gorman et al [4], confirming that under this system the workflow among professional groups is indeed conceptualized as linear, stepwise and unidirectional; the flow runs mainly from physicians to the other professional groups. In our case, the system caused the physicians to dominate other groups, whose work became contingent on the timely and appropriate execution of physicians' tasks. In order to fairly distribute the benefits of work efficiency, these systems should support real-time, ad hoc, and mutual relationships in the medication process. Nurses' and pharmacists' inputs into this process should also be considered. Only then can the system support actual inter-professional relationships in the medication process.

Acknowledgement

We gratefully acknowledge our interviewees. The first and the second authors were supported by training grants from Urmia University of Medical Sciences (Urmia, Iran).

References

- [1] Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE, et al. Role of computerized physician order entry systems in facilitating medication errors. *Jama*. 2005 Mar 9;293(10):1197-203.
- [2] Aarts J, Berg M. Same systems, different outcomes--comparing the implementation of computerized physician order entry in two Dutch hospitals. *Methods Inf Med*. 2006;45(1):53-61.
- [3] Cheng CH, Goldstein MK, Geller E, Levitt RE. The Effects of CPOE on ICU workflow: an observational study. *AMIA Annu Symp Proc*. 2003:150-4.
- [4] Gorman PN, Lavelle MB, Ash JS. Order creation and communication in healthcare. *Methods Inf Med*. 2003;42(4):376-84.
- [5] Aarts J, Ash J, Berg M. Extending the understanding of computerized physician order entry: Implications for professional collaboration, workflow and quality of care. *Int J Med Inform*. 2007 Jun;76 Suppl 1:4-13.
- [6] Wears RL, Berg M. Computer technology and clinical work: still waiting for Godot. *Jama*. 2005 Mar 9;293(10):1261-3.
- [7] Strauss A. Work and the Division of Labor. *The Sociological Quarterly*. 1985;26(1):1-19.
- [8] Reddy MC, Dourish P, Pratt W. Coordinating heterogeneous work: information and representation in medical care. *Proc of the European Conference on CSCW*; 2001; Dordrecht: Kluwer Academic Publishers; 2001. p. 239-258.
- [9] Bardram J. Designing for the Dynamics of Cooperative Work Activities. *Proc of the 1998 ACM conference on CSCW*; 1998; Seattle, United States: ACM Press, NY, USA; 1998. p. 89 - 98.
- [10] Beuscart-Zephir MC, Pelayo S, Anceaux F, Meaux JJ, Degroisse M, Degoulet P. Impact of CPOE on doctor-nurse cooperation for the medication ordering and administration process. *Int J Med Inform*. 2005;74(7-8):629-41.
- [11] Weir C, Hoffman J, Nebeker JR, Hurdle JF. Nurse's role in tracking adverse drug events: the impact of provider order entry. *Nurs Adm Q*. 2005;29(1):39-44.

Usability Studies on Interactive Health Information Systems; Where Do We Stand?

Linda W.P. PEUTE^a, Richard SPITHOVEN^a, Piet J.M. BAKKER^b Monique W.M. JASPERS^a

^a*Department of Medical Informatics, Academic Medical Center, University of Amsterdam, the Netherlands*

^b*Department Medical Oncology, Academic Medical Center, Amsterdam*

Abstract. This paper discusses the preliminary results of a systematic review of the literature on applied usability studies of health information systems in the period 1990 to 2006. Abstracts were included when they described an evaluation of the usability of a health information system. To gain insight into usability methods applied and their properties we constructed a framework to analyze the studies. The framework includes objectives, designs, number of participants, user-profiles, settings, medical domain, and type of health information systems evaluated. Fifty-two Papers were included in the review. Findings show that from 2002 an increasing trend can be observed of publication of usability studies. Most studies discuss summative usability results on working systems thereby focusing on systems' adoption problems. Formative usability studies lack a uniform way to describe how study results contributed to the system's iterative development cycle.

Keywords. User Interfaces, Assessment-Evaluation, Health Professional Workstation

Introduction

In healthcare, clinicians manage sensitive and complex information while working in a highly agile work environment. A critical prerequisite for computer systems to be successfully implemented in such settings is that their interactive user interfaces are streamlined to the working practices of their users and are highly usable [1]. To verify and optimize system usability, a variety of analytical and empirical methods from the area of usability engineering and human-computer-interaction have been adapted to and applied in healthcare system evaluation studies [2, 3]. It is argued that applying these methodologies to healthcare information systems design and evaluation will lead to an understanding of clinicians' reasoning and processing of health care concepts crucial in system re-design efforts [4]. The diversity in usability methods and type of health care system to which these methodologies have been applied has made it difficult to gain a clear overview on what insights on healthcare systems development have been acquired and where challenges for future usability studies remain. A detailed

investigation of published usability studies may reveal the benefits and trade-offs of usability methods applied in the healthcare environment and give insight into how information needs of targeted health care users may or may not be reached. In this paper we present the preliminary results of a systematic review on applied usability studies on healthcare information systems. The aim is to provide clear insight into what methodologies and study designs have been used in usability studies on health information system development and evaluation. Further research will focus on measured outcomes in usability studies and insights gained into the merits of applied usability studies to health information system development and implementation.

1. Research methodology

1.1. Background definitions

All definitions of computer system *usability* speak of multiple components [4,5,6]. As Nielsen states, it is important to realize that usability is not a single, one-dimensional property of a computer user interface alone [5]. Usability answers the question on how well users can use an information system's functionality, concerning learnability, efficiency of use, memorability, errors and satisfaction. Kushniruk and Patel define usability as the capacity of a system to allow users to carry out their tasks safely, effectively, efficiently and enjoyably. In this review we use both definitions for overall coverage of studies describing a usability evaluation study [6].

For the definition of *health information systems*, the authors adopted the definition of Ammenwerth and de Keizer: 'all computer-based components which are used to enter, store, process, communicate, and present health related or patient related information and which are used by health care professionals or the patient themselves in the context of inpatient or outpatient patient care' [7]. This definition includes laboratory information systems, decision support systems, hospital information systems, management information systems, medical record systems, online systems, radiology information systems and reminder systems. In this review we however focus on interactive health information systems with a computer screen; human-computer interaction is possible by a user-interface. Community networks, databases, MEDLARS and geographical information systems were excluded from the review.

1.2. Selection procedure

MEDLINE and EMBASE were systematically searched on April 21, 2005 and May 2, 2006 using a combination of the following Medical Subject Headings (MeSH terms) *information system*, *evaluation studies* and the keywords directly associated with the definition of "usability". Excluded MeSH terms *Community networks*, *databases*, *MEDLARS*, *geographical information systems*. Additional searches with the same keywords were performed by searching ScienceDirect and Proceedings of AMIA and IEEE. Reference lists of articles were also reviewed for relevant articles. Dutch and English publications, published between January 1990 and February 2006, were included. This resulted in 212 articles. To gain insight into the general features and quality of the published usability studies on health information systems between 1990 en May 2006 article inclusion was based on abstracts. The abstracts were reviewed and classified (inclusion, exclusion, unclear) independently by the first and second author.

Studies were included if in the abstract mention was made of (1) a computer system that complied with the stated description of interactive health information system and (2) an evaluation study in which aspects of system usability were assessed. After reviewing the abstracts 52 articles met the inclusion criteria [8]. Inclusion of more than one article on a single system was done if the articles described distinct evaluations. The inter-rater agreement between reviewers on the eligibility of studies for the systematic review was $k=0.81$. All disagreements were resolved by reviewing and discussing the complete articles ($n=19$) before deciding to include or exclude them.

1.3. Framework for article review

Each included article was first rated on general characteristics: year/ month of publication, authors and source/ journal. To analyze the described usability studies a framework was developed by the authors based on definitions from the field of evaluation studies and terms applicable to usability studies as described by Patel, Kushniruk, Nielsen, and Wyatt [4,5,6,9]. The framework includes the following dimensions (1) study objective(s): *formative* (evaluation is done parallel to system development and implementation, results are fed back to system design) / *summative*, (evaluation is done after system introduction is completed; aim is to check whether aims of system implementation are fulfilled), *subjectivist* (*mostly qualitative approach*) / *objectivist* (*mostly quantitative approach*); (2) study design (*within group/ single group/ between group*); (3) setting (*laboratory, in practice, field study*); (4) methods (*observation, questionnaire, survey, interviews, logging, GOMS, pluralistic walkthrough, cognitive walkthrough, guideline checklists, screenshot analysis, heuristic evaluation, cognitive task analysis, additionally mentioned methods*), (5) outcome measurements, (6) number of participants, (7) user-profiles (*subjects computer expertise, subjects expertise in work domain, subjects role in workplace*), (8) role of researcher (*neutral agent, system developer, add. .*); (9) medical domain and (10) type of health information systems described in the studies.

The two reviewers independently extracted the information from all included articles. The raw agreement of extracted information was 89%. Disagreements were 100% resolved by consensus. Extracted information was then entered in a Microsoft Access© database specifically constructed by the authors, based on the framework, to analyze the data.

2. Results

2.1. General characteristics of included studies

Of the total of 52 included articles only 14 (39 %) were published before January 2002. 19 (37%) of the selected studies were drawn from Conference Proceedings in 2002 and 2004. Of the 52 studies 7 (13%) were set as 'field study', 31 (60%) 'in practice', 5 (10%) 'laboratory study', and in 9 studies (17 %) the specific study setting was not mentioned, nor could be deduced. Systems evaluated were web-based systems (8%), PACS (6%), electronic patient records (6%), electronic medical record systems (8%), computerized physician order entry systems (8%), decision support system (8%),

Table 1. Overview of formative and summative usability studies and frequency of methods applied

Objective/ design	No. studies	Obs	Int	Log	Sur	GoMs	CW	CTA	TA	Heu
Form./ Subj.	8	2	1	1	6	1	4	1	6	3
Form./ Subj. → Obj.	6	1	3	-	5	-	1	1	1	1
Form./ Obj.	0	-	-	-	-	-	-	-	-	-
Form./ Obj. → Subj.	0	-	-	-	-	-	-	-	-	-
Summ./ Subj.	18	5	2	4	8	1	-	1	5	4
Summ./ Subj. → Obj.	15	6	4	6	10	-	-	-	3	5
Summ./ Obj.	3	1	-	1	2	1	-	-	-	-
Summ./ Obj. → Subj.	2	-	-	-	2	-	-	-	-	-
Totals	52	15	10	12	33	3	5	3	15	13

Key: Obs - observation of real life working situations, Int - interviews, Log - logging, Sur - Survey/questionnaire, GoMs - Goals, Operators, Methods, and Selection Rules, CW - Cognitive Walkthrough, CTA - Cognitive Task analysis, TA - Think Aloud, Heu - heuristic analysis, Form: formative study, Summ: Summative study, Obj: objective, Subj: subjective.

of which the majority of systems evaluated were specialized systems such as clinical reminder systems, and dental systems. Of the described usability studies concerning health information system evaluation, 34 (65%) used a single group design, 9 (17%) used a between group study design, 1 (2%) used a within group study design, and 8 (15%) only mentioned the methods applied and results without an explicit description of the study design. Studies were divided in formative studies (27%) and summative studies (73%). Table 1 specifies the studies' objective in potential study design combinations; formative/ summative studies with a subjectivist (mostly qualitative) or objectivist (mainly quantitative) approach. If multiple usability methods were applied with a differing study design, this is shown by an arrow →. If the first usability analysis was more qualitative and the second analysis more quantitative then this is shown by subj. → Obj., and vice versa.

2.2. . Applied methodology in the usability study description

In only 8 (15%) of the studies, a singular evaluation method was used, of which 6 applied a usability survey aimed at measuring end-user satisfaction with the system. Cognitive analysis methods such as the Think Aloud (TA), and Cognitive Task analysis (CTA) were combined with a heuristic analysis (Heu) (based on usability guidelines), logging (Log) (log file analysis) and use in real practice by observation of real life working situations (Obs) and (semi-structured) interviews (Int). All methods with exception of the 'Cognitive Walkthrough method (CW) have been applied in both formative as well as summative studies. In non of the selected studies the usability methods: 'pluralistic walkthrough', 'screenshot analysis' and 'usability guidelines' had been applied as a usability data collection method. In 8 studies 'focus groups', 'contextual inquiry', 'expert evaluation' and 'activity analysis', and 'remote web-based usability testing' were mentioned as complementary usability methods applied in the

Table 2. Percentage of selected studies per system design phase in iterative system engineering.

System development phase	Perc. Usability studies	Applied Methods
Requirements phase	4%	Think Aloud , Interviews, User-requirements questionnaire
Specification/ development phase	19%	Heuristics, Cognitive Task Analysis, GOMMS
Implementation/ release phase	0%	-
Post-implementation analysis for insight into the need for system re-engineering or end-user satisfaction	73%	Surveys/ developed questionnaires, Cognitive Walkthrough, Logging and Observation of work practices, Heuristics, Think Aloud
Additional: system selection of commercially build system	4%	Comparative methods; survey/ requirement questionnaire

system design phase. Most studies (73%) perform usability evaluations on working systems, in order to detect system errors or analyze usability issues related to end-user satisfaction with system use. Table 2 shows the percentage of studies applied in the 5 stages of iterative system development. Only 19% of the studies actually apply usability methods within the specification/ development phase. Non of these studies accurately described the collected results and how the results of the usability study were applied in system optimization or re-engineering.

3. Discussion

In published literature it has come to the fore that the ultimate acceptance or rejection of health information systems largely depends on the degree of system usability [10,11]. From 2002 an increasing trend can be observed of publications on usability studies of health information systems which indicates that usability is gaining more notice in evaluation research. It appears that usability evaluation studies on healthcare information systems indeed mostly aim at gaining insight into system aspects that influenced its adoption as 60% of the performed usability studies take place ‘in practice’ and 73% describe summative results. Though it is also widely acknowledged that usability studies are crucial in designing user interfaces of health information systems that fulfil clinicians’ needs [11] only 23% percent of the usability studies were performed as part of a system development cycle. It is clear that a qualitative usability methodology is time-consuming and integrating its results in a system’s iterative development cycle is difficult. However, these results seem to point out that usability experts are called in too late in a system development process. Use of healthcare information systems is very sensitive in regard to health hazards of patients. The additional strain placed on clinicians by a new system may be a nuisance, but more importantly it can lead to dangerous errors. If usability experts are included from the beginning of the development process not only the iterative design cycle may shorten, more importantly new insights might be acquired on general system design aspects that might potentially lead to errors in healthcare. Until now, only few studies (3%) actually relate system usability issues to the potential of errors in healthcare caused by system use [3].

The rapid and continuous developments of commercially based health information systems make it even more important that these systems should be safe, with clear and

validated studies on their usability, preferable based on scientific insights into usability and general metrics on system safety aspects.

Next to this, the performed evaluation studies which describe the assessment of system specific usability aspects lack a uniform description of the study and of its results. As most studies offer practical insights into system adoption aspects the potential merits of using the various methods are still unclear. For example, when does the Think Aloud or Cognitive Walkthrough method offer the best results? The theory of triangulation [12] proposes that methods and approaches should be combined for better validation of study results. Of all analyzed studies, 23% combined two or more qualitative methods and 44% combined qualitative usability methods with survey and interview analysis. However, a uniform description of usability study results and applied methods is needed in order to gain insights into general usability guidelines for healthcare system development.

4. Conclusion

Integration of usability in the development processes of health care IS is challenging. Insights into where usability has effectively been integrated in design and evaluation may lead to the development of new metrics on which to evaluate healthcare user interfaces. Future analysis of studies of the systematic review will focus among other things on the applied strategies of usability methods on different types of healthcare information system and the experiences and lessons learned of combined methodologies in usability evaluation studies.

References

- [1] Ash J, Coiera A, Berg M. Some unintended consequences of information technology in health care: the nature of patient care information systems. *J Am Med Inform Assoc* **11** (2004), 104-112
- [2] P.C. Tang, V.L. Patel, Major issues in user interface design for health professional workstations: summary and recommendations, *Int J Biomed Comput* **34** (1994), 139-48
- [3] Kushniruk AW, Triola MM, Borycki EM, Stein B, Kannry JL. Technology induced error and usability: The relationship between usability problems and prescription errors when using a handheld application. *Int J Med Inform* **74** (2005), 519-526.
- [4] V.L. Patel, A.W. Kushniruk, Interface design for health care environments: the role of cognitive science, *Proc. AMIA Symp* (1998), 29-37.
- [5] Nielsen J. *Usability engineering*. Cambridge 1993, MA: Academic Press.
- [6] ISO 9241-11, *Guidance on usability*, 1998.
- [7] E.Ammenwerth, N. de Keizer, An inventory of evaluation studies of information technology in health care: Trends in Evaluation research 1982 – 2002, *Methods Inf Med* **44** (2005), 44-56.
- [8] L.W.P.Peute, M.W.M.Jaspers. *Usability Studies of Healthcare Information Systems; Systematic Review*. Technical report TR2008, Dept. of Medical Informatics, Academic Medical Center, University of Amsterdam, 2008. Available at <http://kik.amc.uva.nl/KIK/reports>.
- [9] Friedman CP, Wyatt JC. *Evaluation methods in medical informatics*. New York: Springer; 1997
- [10] H. Laerum, G. Ellingsen, A. Faxvaag, Doctors' use of electronic medical records systems in hospitals: cross sectional survey, *BMJ* **323** (2005),1344-8.
- [11] A. Kushniruk, Evaluation in the design of health information systems: application of approaches emerging from usability engineering, *Comput Biol Med* **32** (2002), 141-9.
- [12] E. Ammenwerth, S. Gräber, G. Herrmann, T. Bürkle, J.A. König, Evaluation of health information systems-problems and challenges, *Int J Med Inform* **71** (2003), 125-35.

Combination of Short- and Longaxis MR Image Sequences for the 3D Segmentation of the Left Ventricle

Dennis SÄRING^{a,1}, Jatin RELAN^a, Michael GROTH^b, Kai MÜLLERLEILE^c and Heinz HANDELS^a

^a *Department of Medical Informatics*

^b *Department of Diagnostic and Interventional Radiology*

^c *Department of Cardiology/Angiology*

University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany

Abstract. Segmentation of the left ventricle (LV) is required to quantify LV remodeling after myocardial infarction. Therefore spatiotemporal Cine MR sequences including longaxis and shortaxis images are acquired. In this paper a new segmentation method for fast and robust segmentation of the left ventricle in 4D MR images is presented. The new approach considers the position of the mitral valve and the apex as well as the longaxis contours to generate a 3D LV surface model. The segmentation result can be checked and adjusted in the shortaxis images. Finally quantitative parameters were extracted. For evaluation the LV was segmented in eight datasets of the same subject by two medical experts using a contour drawing tool and the new segmentation tool. The results of both methods were compared concerning robustness and interaction time and intra- and interobserver differences. The presented segmentation method proved to be fast and robust and the intra- and interobserver differences are decreased significantly.

Keywords. Cardiovascular magnetic resonance, Image processing, Left ventricular segmentation

Introduction

Cardiovascular disease is reported as the number one cause of death globally. Approximately 280,000 Germans suffer from a myocardial infarction each year. The infarcted area of the myocardium loses its ability to contract and the remaining healthy muscle needs to compensate for that weakened area. This yields to left ventricle (LV) remodeling which is characterized by e.g. decreased LV ejection fraction. The quantification of LV remodeling based on volume and mass parameters is an indicator for diagnosis and treatment planning. In 2006, Säring et al. presented the software

¹ Corresponding Author: Dennis Säring, University Medical Center Hamburg-Eppendorf / Department of Medical Informatics, Martinistraße 52 (Bldg. S14), 20246 Hamburg, Germany; E-mail: d.saering@uke.uni-hamburg.de.

system HeAT (Heart Analysis Tool) for the quantitative analysis of 4D MR image sequences [1]. The analysis was based on a manual LV segmentation where the user

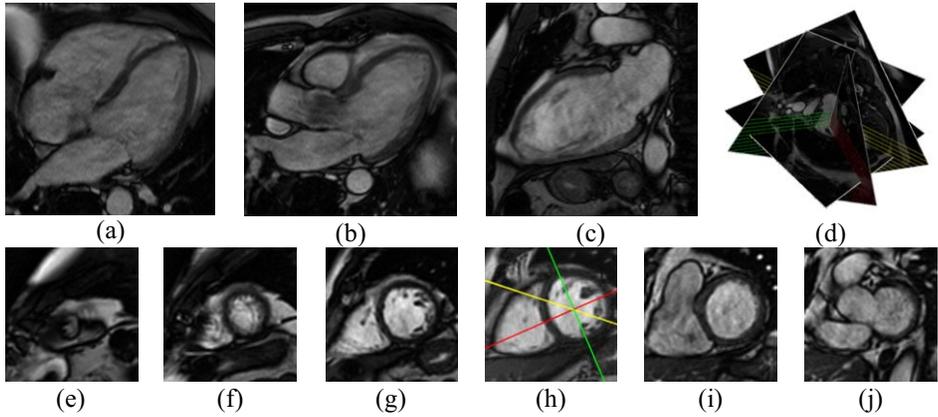


Figure 1. 4D Cine MR sequence: 2D images in 4-chamber longaxis (a), 3-chamber longaxis (b), 2-chamber longaxis (c) and a sequence of shortaxis (e-j) views from the apex to the mitral valve. Position and orientation of 4-chamber (red), 2-chamber (green) and 3-chamber (yellow) image in a midventricular SA slice (h) and in 3D coordinate system (d).

has to define endocardial and epicardial structures in all shortaxis images. This segmentation process is time-consuming and observer-dependent.

In this paper a new semi-automatic segmentation method for the extraction of the myocardium of the left ventricle combining longaxis (LA) and shortaxis (SA) images is presented. Recently several approaches for the LV segmentation were published. These approaches can be divided into intensity, shape and model based methods [2,3]. Generally these methods operate on shortaxis images (Fig. 1(e-j)) only. Due to the slice thickness of 10mm and the orientation of SA images in clinical Cine MR datasets the definition of the ending (mitral valve) slice is ambiguous, because of different numbers of SA images has a strong influence to the computed volume and mass parameters. Therefore longaxis images in 4-chamber, 3-chamber and 2-chamber view have to be included to identify the position of the mitral valve plane in 3D space (Fig. 1(a+c)). Goshtaby et al. [4] show the fusing of LA and SA images into an isotropic volume image using voxel interpolation without posterior LV segmentation. In van Geuns et al. [5] SA contours are generated based on LA contours using spline interpolation. Here the global parameters are computed using the Simpson rule on the number of SA contours without including information concerning the mitral valve position.

1. Materials and Methods

The goal of this work is to develop a fast and robust segmentation method with reduced inter- and intraobserver variability. Therefore LA information about position and orientation of the mitral valve plane (*MVP*) and the apex as well as the myocardial contours (endocardium, epicardium) are used to generate a 3D surface model of the left ventricle. Due to the fact that only 6 LA contours have to be outlined in our approach the time for the segmentation process will be decreased. After generation the shape of the model can be visually inspected and if needed manually adjusted in SA images.

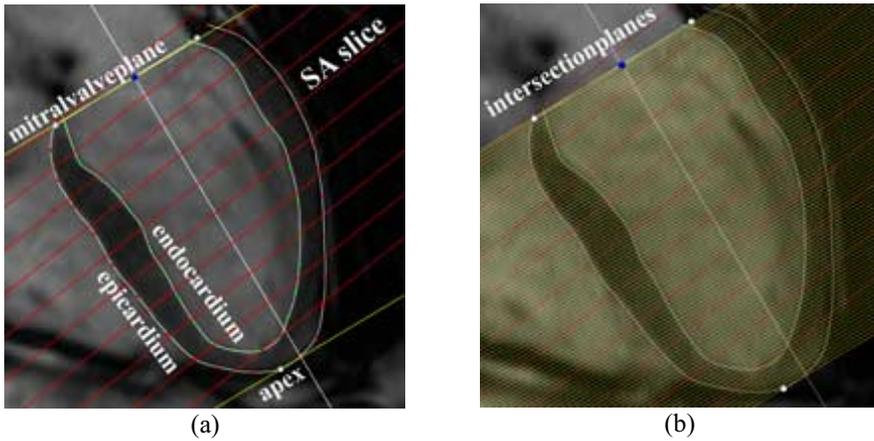


Figure 2. Position of the mitral valve plane and SA images (a) and the intersection planes covering the whole LV with a slicethickness of 1mm (b)

1.1. Material

Unfortunately no gold standard for LV segmentation exists and software phantoms are often used to evaluate new methods [6]. In this paper four datasets ($D_{1,2,3,4}$) from the same subject but in different subject positions are acquired. The slice thickness s is changed in comparison to the clinical standard ($D_{1,2}$: $s = 6\text{mm}$, no gap; D_3 : $s = 8\text{mm}$, gap = 2mm; D_4 : $s = 10\text{mm}$, gap = 2mm; standard: $s = 8\text{mm}$, gap = 2mm). Two of these datasets D_k with $k = 1; 2$ were divided into subdatasets with different numbers of slices, different start slices and different s .

- D_k^1 is built by appending all odd numbered slices of D_k
- D_k^2 is built by appending all even numbered slices of D_k

In the ideal case the calculated volume and mass parameters for all datasets should be equal. $D_{1,2,3,4}$ are acquired in 4-chamber, 3-chamber and 2-chamber longaxis view (LA) and a sequence of shortaxis (SA) views.

1.2. Methods

In this paper LA and SA images are transferred from 2D in a 3D coordinate system based on the DICOM header information solving a linear equation [4] (Fig. 1(d)).

The mitral valve and the apex as well as endocardial and epicardial contours are outlined in the LA 2D views (Fig. 2(a)). The mitral valve plane (MVP) is built in 3D using position information of both LA views. Then intersection planes (ISP_i with $i = 1, \dots, n$) with a distance of 1mm are generated automatically parallel to the MVP covering the whole left ventricle from the apex to the mitral valve (Fig. 2(b)).

For each plane ISP_i the six endocardial $endo_i^j$ and six epicardial epi_i^j ($j = 1, \dots, 6$) intersection points are calculated (Fig. 2(c)). Defining these intersection points as seedpoints enables the generation of smooth in-plane contours C_i^{endo} and C_i^{epi} using Bezier interpolation. 3D surface models for endocardium and epicardium are constructed based on LA contours and all in-plane contours C_i^{endo} and C_i^{epi} for $i = 1, \dots, n$. Additionally contours in the original shortaxis views are extracted by cutting the surface models in 3D space

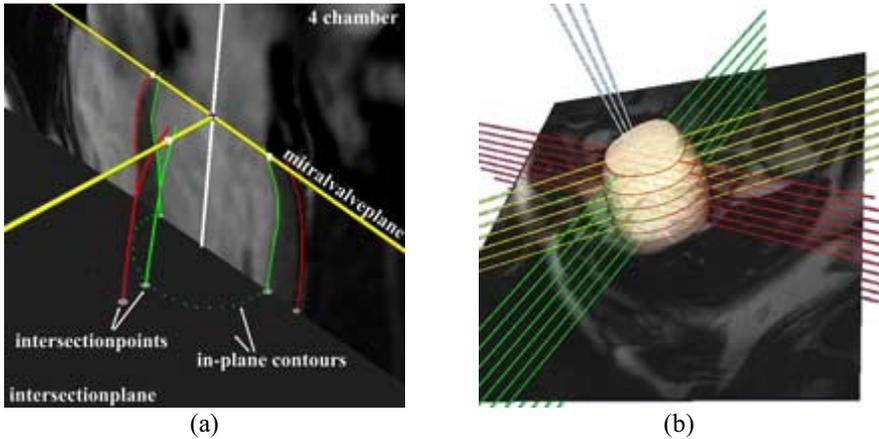


Figure 3. Generation of the in-plane contours based on the intersection points (a) and visualisation of the cutted SA contours. Intersection lines of the SA sequence with the LA images are displayed (b); 4-chamber view: red, 3-chamber view: yellow, 2-chamber view: green

with the image plane. These contours can be reviewed and adjusted. Mass and volume parameters are calculated using the adjusted 3D surface models.

1.3. Evaluation

For evaluation purpose endocardial and epicardial contours of the datasets $D_{1,2,3,4}$ in end-diastolic and end-systolic phase in all shortaxis MR images are outlined by two medical experts (obs_1, obs_2) using a contour drawing tool [1]. Each observer traced the contours of each dataset two times with the manual segmentation. Therefore approx. 500 contours have to be defined. Based on these contours five global parameters (e.g. enddiastolic volume (EDV), LV ejection fraction (LVEF)) were extracted for each dataset using the Simpson rule. Then the presented segmentation method is used to compute the same global parameters including LA information. For each dataset only 6 LA contours need to be defined. If needed the shape of the generated model is manually adjusted. Then two comparison sets were obtained, each based on the eight datasets $D_{1,2,3,4}$ and $D^{1,2}_k$ with $k = 1,2$: measurements obtained by obs_1 the first time versus the second time (Intraobserver variability) and the averaged difference of measurements obtained by observer obs_1 the first time versus obs_2 and those obtained by obs_1 the second time versus obs_2 (interobserver variability). Furthermore the influence of the different slice thickness ($s = 6; 8; 10; 12\text{mm}$) of the MR datasets of the same subject to the extracted parameters

2. Results

Four 4D Cine MR image sequences were obtained at the Department of Diagnostic and Interventional Radiology from one subject in different positions using a Siemens 1.5T MR scanner. Longaxis and shortaxis images were acquired by using

electrocardiographically triggered breath-hold imaging techniques according to the American Heart Association (AHA) scientific statement. For each slice 20 phases

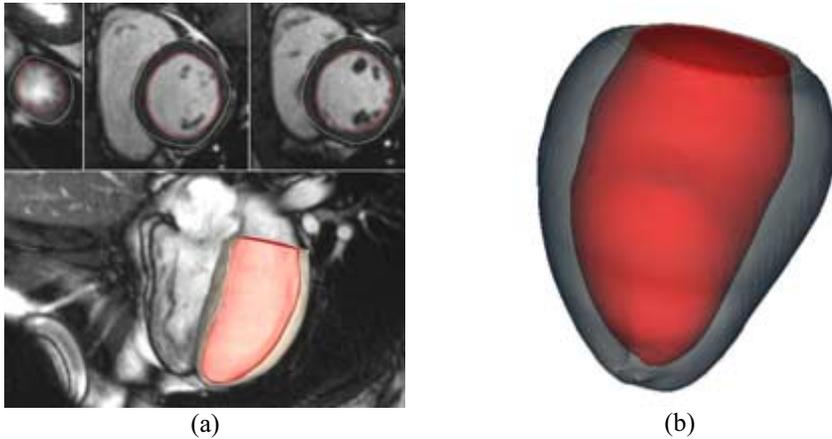


Figure 4. Result of the segmentation process: Contours of the surface overlaid in SA slices and 4-chamber view (a). Final surface model of the left ventricle in diastolic phase (b)

provided the complete coverage of the cardiac cycle resulting in a total of 380 cardiac images per sequence. The inter- and intraobserver differences comparison between global LV function for manually and semi-automatically segmented images are shown in Tab. 1+2. The mean and standard deviation of the differences in [ml] and [%] of all parameters is decreased. In case of intraobserver comparison e.g. the mean EDV difference between first time and second time of obs_1 is reduced from 13.7% to 2.3%. Concerning slice thickness the mean differences of EDV (14.1% to 5.8%) and ESV (18.8% to 4.9%) are decreased but SV (10.7% to 20.7%) and LVEF (4.3% to 11.6%) are increased.

Table 1. Intraobserver difference comparison of obs_1 between global LV function for manually and semi automatically segmented images [mean difference \pm std]. (EDV: end-diastolic volume; ESV: end-systolic volume; SV: stroke volume; LVEF: left ventricular ejection fraction; Mass: mass of the myocardium)

	manual		semi-automatic	
	absolute	relative	absolute	relative
EDV [ml]	6.4 \pm 0.5	4.5 \pm 0.8	1.7 \pm 0.1	1.3 \pm 0.0
ESV [ml]	5.9 \pm 7.5	8.8 \pm 11.2	0.3 \pm 0.2	0.5 \pm 0.3
SV [ml]	5.5 \pm 1.6	7.3 \pm 1.1	2.9 \pm 0.6	4.0 \pm 0.3
LVEF [%]	4.2 \pm 2.7	7.6 \pm 4.5	1.5 \pm 2.6	0.8 \pm 1.2
Mass [g]	7.8 \pm 5.3	7.5 \pm 4.7	6.2 \pm 6.1	4.8 \pm 4.7

3. Discussion

A semi-automatic LV segmentation method has been developed that combines LA information and intersection planes as well as SA images. In the presented work global parameters including position and orientation of the mitral valve and the apex are extracted. The generation of intersection planes and in-plane contours covering the

whole left ventricle enable the construction of a surface model (Fig. 3(a+b)) and its quantitative analysis. The inter- and intraobserver differences can be decreased for all extracted parameters. Using the presented approach the time needed for the segmentation process is decreased from 60 minutes to 5 minutes including manual adjustment. So the semi-automatic segmentation using LA information proved to be fast and robust for the quantification of LV mass and volume properties. Currently the presented segmentation tool is used to evaluate LV remodeling based on 4D Cine MR

Table 2. Interobserver difference ($obs_1 - obs_2$) comparison between global LV function for manually and semi-automatically segmented images [mean difference \pm std]

	manual		semi-automatic	
	absolute	relative	absolute	relative
EDV [ml]	17.8 \pm 11.7	13.7 \pm 9.5	2.9 \pm 2.1	2.3 \pm 1.7
ESV [ml]	7.9 \pm 8.0	11.3 \pm 11.6	3.1 \pm 2.0	4.9 \pm 3.1
SV [ml]	13.0 \pm 13.4	22.7 \pm 25.9	5.6 \pm 6.1	9.1 \pm 2.6
LVEF [%]	9.4 \pm 2.7	20.9 \pm 8.2	3.1 \pm 4.1	6.3 \pm 8.4
Mass [g]	15.5 \pm 13.7	6.3 \pm 5.0	1.1 \pm 0.8	0.6 \pm 0.5

datasets of thirty patients with myocardial infarction in a baseline (acute infarction phase) vs. follow-up (approx. 15 month later) study. In future the variability concerning slice thickness has to be evaluated. Also the influence of papillary muscles and trabecular structures has to be analyzed to improve the quality of the extracted global parameters. Automatic adaptation of the contours to all time points will enable the generation of a 4D heart model to visualize dysfunctional areas of the left ventricle.

References

- [1] D. Säring and et al. Computer-assisted analysis of 4d cardiac mr image sequences after myocardial infarction. *Methods Inf Med*, 45(4):377–383, 2006.
- [2] D. Rueckert and et al. Shape-based segmentation and tracking in 4d cardiac MR images. In *CVRMed*, pages 43–52, 1997.
- [3] L.J. Spreeuwiers and et al. Myocardial boundary extraction using coupled active contours. In *Computers in Cardiology, 2003*, pages 745–748, 21-24 Sept. 2003.
- [4] A. A. Goshtasby and et al. Fusion of short-axis and long-axis cardiac mr images. *Comput Med Imaging Graph*, 20(2):77–87, 1996.
- [5] Robert Jan M van Geuns and et al. Automatic quantitative left ventricular analysis of cine mr images by using three-dimensional information for contour detection. *Radiology*, 240(1):215–221, Jul 2006.
- [6] W. P. Segars and et al. A realistic spline-based dynamic heart phantom. *Nuclear Science, IEEE Transactions on*, 46(3):503–506, 1999.

An automated method for analyzing adherence to therapeutic guidelines: Application in Diabetes

Massoud TOUSSI, MD, MSc ^{a,1}, Vahid EBRAHIMINIA, MD, PhD ^a, Philippe Le TOUMELIN, MD ^a, Régis COHEN, MD ^b, and Alain VENOT, MD, PhD ^a

^a *Lim&Bio Laboratoire d'Informatique Médicale et de Bioinformatique, University of Paris Nord,*

^b *Department of Endocrinology,*

Avicenne Hospital, Assistance Publique-Hôpitaux de Paris, France

Abstract. Background: Physicians' adherence to guidelines can be used for measuring prescribing appropriateness. We present a simple approach allowing the automation of this process. Design: The drug therapy is described in terms of treatment type, pharmacotherapeutic classes, international non proprietary names (INN) and doses. A rule-based engine implementing the guideline generates recommendations for each patient record. These are automatically compared with prescriptions of the same patient in three levels of detail. Participants: Ambulatory patients admitted for the follow-up of their type 2 diabetes between June 2003 and September 2004 in a university hospital in France. Results: For 574 patient records included in the study, physicians agreed with the guideline recommendations over the choice of type of treatment in 473 cases (82%). When agreement over pharmacotherapeutic class of drugs was also taken into account, the adherence ratio decreased to 448 cases (78%). Finally, when the dosage of each drug was taken into account, the adherence ratio dropped to 396 cases (69%). Adherence ratios were also dependent on the type of treatment at admission: low for patients on oral tritherapy, and on diet and exercise. The results also highlighted inertia of physicians for beginning drug therapy and the underuse of biguanides. Conclusions: The proposed method provides an automatable way of measuring the appropriateness of treatment choice, which can be used for chronic diseases.

Keywords: assessment-evaluation, compliance, quality management, data analysis-extraction, evidence based guidelines, decision support, prescribing appropriateness, diabetes mellitus, therapeutic strategy.

Introduction

Assessing the quality of drug prescriptions is an important issue for which indicators of the appropriateness have been developed [1,2,3]. Some of these indicators are based on

¹ Corresponding Author : Lim&Bio Laboratoire d'Informatique Médicale et de Bioinformatique, Faculté de Médecine Léonarde de Vinci, Université de Paris 13, 74 rue Marcel Cachin, 93017 Bobigny, France ; Email : massoudtoussi@gmail.com.

the contents of summaries of product characteristics (SmPC). They can be used to assess the safety of prescriptions, based on the number of contraindications, drug interactions, dosage problems, etc. However, the assessment of the appropriateness of treatment choice goes beyond the SmPCs and requires comparison with recommendations from clinical guidelines [4]. For instance, antihypertensive drugs of various pharmacotherapeutic classes may have the same indication (hypertension) in SmPCs, but the choice of a specific class is generally based on the knowledge found in clinical guidelines (e.g., ACE inhibitors they may be preferred in diabetic patients).

Numerous studies have shown controversial results regarding adherence to guidelines even for a single disease [5,6,7], and the overall adherence is known to be mediocre [8,9]. It remains unclear whether the differences in adherence ratios result from differences in practices, in methods used to measure adherence, or both [10,11].

In this article, we present a simple method for quantifying the adherence to a therapeutic guideline and for easily analyzing the situations in which recommendations are not followed by physicians. This method is applied to type 2 diabetic patient data.

1. Materials and Methods

1.1. The description of treatments

We described the therapeutic prescriptions of physicians in terms of “type of treatment”. For example, the types of treatment for diabetes include diet and exercise (i.e. zero drug therapy), oral monotherapy, bitherapy, tritherapy, and insulin therapy. Each drug in a prescription is linked to some “pharmacotherapeutic class”, “international non proprietary name (INN)”, and “dose” (see Figure 1).

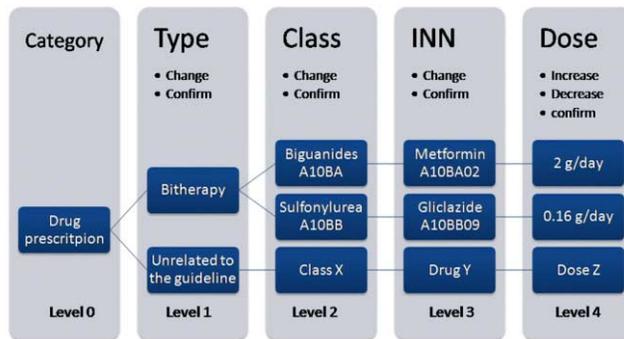


Figure 1. The four levels of detail in a drug prescription

1.2. French clinical guidelines for the management of type 2 diabetes

We used the official guideline of the French National Authority for Health (FNAH) for the management of type 2 diabetes [12]. It was the national reference standard evidence-based guideline at the time of the study, and was widely available to physicians in electronic and paper forms. The guideline proposes a step-by-step therapeutic strategy by dedicating a section for each type of treatment.

Recommendations are provided by their grades of evidence. The goal of treatment is to maintain Hb_{A1C} inferior or equal to 6.5%. If this goal is not obtained within a step, the guideline recommends increasing the dose; if not possible: replacing the drug with one from a more efficient class; and if this is not possible: passing to the subsequent type of treatment. Specific recommendations are given for the obese, the elderly, and the patients with renal insufficiency.

1.3. Patient data

We used a database of electronic records of ambulatory patients admitted for the follow up of their type 2 diabetes to the endocrinology department of Avicenne university hospital of Bobigny, France. The patients came to the hospital for periodic control of their diabetes or because of its deregulation. They passed a few hours in the hospital during which they had a laboratory exam and a visit by a physician. During the consultations, physicians did not have access to patient specific recommendations but they could use the guideline as a whole in paper form. All admissions from June 2003 to September 2004 were included. The database contains demographic, anthropometric, clinical, laboratory and therapeutic information. Parameters used by the guideline include age, body mass index, creatinine clearance, ketonuria, and therapeutic history.

1.4. Measuring the adherence

We checked firstly if the type of treatment prescribed by the physician was the same as the one recommended by the guideline. For those prescriptions that fulfilled this condition, we verified if the pharmacotherapeutic classes of drugs were also the same as recommended by the guideline. Based on the anatomical therapeutic chemical (ATC) classification [13], for type 2 diabetes these classes include biguanides (A10BA), sulfonamides (A10BB and A10BC), alphasglucosidase inhibitors (A10BF), and the insulin family (A10A). Two classes; namely thiazolidinediones (A10BG) and other blood glucose lowering drugs (A10BX) are not represented in the studied version of the guideline. Finally, in prescriptions that conformed to the guideline in both type and classes, we checked whether physicians followed the guideline in making the same drug dose intervals (low, moderate, high). We did not check prescriptions for conformity in INN names because this level is not represented in the studied guideline.

1.5. Computer methods

We developed a computer system in Visual Basic® programming language with four major components (see Figure 2). The first two components analyze and abstract prescription data that are entered by physicians. Another component implements the guideline in a rule-based engine which generates the type of treatment, and class and dose of each drug. Experts verified the results generated by this component for about 25% of cases (selected randomly) in order to be sure of their exactness. The fourth component calculates the agreement between two formalized therapies: one resulting from the abstraction of prescription, and the other generated by the third component. They are calculated in three levels of type, class, and dose. In cases for which the rule-based engine cannot generate a treatment due to a lack of knowledge in the guideline, the comparison engine considers the prescription as conformable.

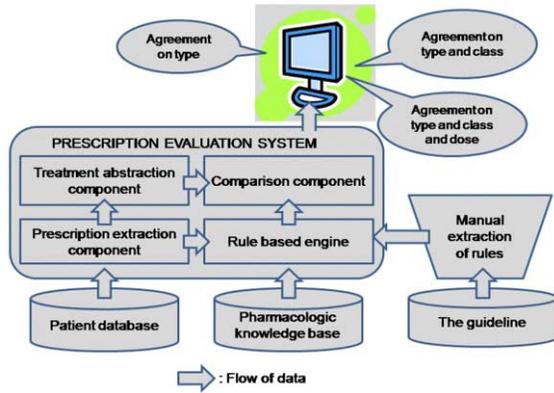


Figure 2. Calculation of agreement values

2. Results

Thirteen patient records containing thiazolidinedions and other glucose lowering drugs were excluded because the guideline lacked recommendations for these classes. A total of 574 patient records were included in the study and were analyzed by the computer system. The mean age of patients was 59.9 (SD = 11.4) years, the mean body mass index was 29.0 (sd = 6.0) kg/m², the mean Hb_{A1C} ratio was 7.9 (SD = 1.7) percent and the mean duration of diabetes since its onset was 11.4 (SD=9.2) years.

2.1. How consistently do physicians follow the guideline?

The type of treatment conformed to the guideline in 473 (82%) cases, with a range of 11 to 96 percent, depending on the type of treatment at admission. These values decrease when the class and dose are also taken into account (see Table 1).

Table 1. Agreement between prescriptions and recommendations for each group of patients with the same type of treatment at admission

Type of treatment at admission	Agreement between prescriptions and recommendations at discharge at three levels			
	Type only n (% in row)	Type and class n (% in row)	Type, class and dose n (% in row)	Number of patients
Diet and exercise	18 (62)	18 (62)	18 (62)	29
Oral monotherapy	127 (86)	117 (79)	109 (74)	148
Oral bitherapy	119 (82)	112 (77)	105 (72)	146
Oral tritherapy	4 (11)	4 (11)	4 (11)	38
Insulin alone or combined with oral therapy	205 (96)	197 (92)	160 (75)	213
Total	473 (82)	448 (78)	396 (69)	574

Table 2. For each group of patients with the same type of treatment at admission, the choices of physician are compared with those of the guideline. Ph: physician's prescription; GL: guideline.

Type of treatment at admission	Prescribed and recommended types of treatment at discharge										Number of patients
	Diet and exercise		Oral monotherapy		Oral bitherapy		Oral tritherapy		Insulin alone or combined with oral therapy		
	Ph	GL	Ph	GL	Ph	GL	Ph	GL	Ph	GL	
Diet and exercise	28	19	1	10	0	0	0	0	0	0	29
Oral monotherapy	1	0	138	134	3	12	0	0	6	2	148
Oral bitherapy	0	0	1	0	126	139	0	0	19	7	146
Oral tritherapy	0	0	0	0	1	0	26	5	11	33	38
Insulin alone or combined with oral therapy	0	0	0	0	1	0	1	0	211	213	213
Total	29	19	140	144	131	151	27	5	247	255	574

2.2. What do physicians do when they do not follow the guideline?

Physicians were less likely to start a monotherapy for patients on diet and exercise. This is shown by the fact that among 29 patients on diet and exercise only one is discharged with monotherapy versus 10 patients for the guideline (see Table 2). Further comparisons showed that physicians were less likely to change the class of a monotherapy from sulfamides or alphaglucoSIDase inhibitors to biguanides.

3. Discussion and conclusion

We proposed a method of measuring the appropriateness of treatment choice by use of a therapeutic guideline. We computerized our method and demonstrated it for a database of type 2 diabetic patients.

Like many other guidelines, our guideline did not contain recommendations for a number of items such as new pharmacotherapeutic classes. Although at the moment of study, HAS released a new version of the guideline, we preferred working with the previous one in order to be sure of physician awareness at the moment of prescriptions. We considered prescriptions conformed to the guideline in situations where the latter does not provide recommendations. Although this convention is quite common in measuring the adherence, it does not literally mean that physician followed the guideline. It would be useful to ask physicians on the reasons of non-adherence. We did not do so because our goal was to analyze the agreement only by use of patient records, in a way that the procedure could be automated.

We used the types of treatments, the pharmacotherapeutic class, and the dose of medications as principal levels of detail in drug therapy and we based our comparisons on these levels. The type of treatment is unspecific to diabetes. In a variety of conditions, such as hypertension and dyslipidemia, patients are treated with non-pharmacological, mono-, bi- and tritherapy. In other conditions, such as asthma, heart

failure and chronic infectious diseases, the therapeutic strategy often comprises several drug regimens, which may be considered as types of treatment. Pharmacotherapeutic class and dose are known concepts in therapeutics and can be applied to all pharmacological treatments.

Existing approaches for measuring adherence to guidelines are often based on physician surveys, patient records, guideline impacts, and multi-level explicit criteria [14,15,16]. Our method seems to be less comprehensive than methods based on explicit criteria, because it does not measure the adherence to all features of guidelines, such as diagnostics. However, it was not among our objectives to propose a comprehensive tool for measuring the adherence to guidelines.

Our method can be used both as a tool for assessing the appropriateness of choice of treatment, and as an educational aid for physicians. We are actually trying to combine this technique with data mining techniques in order to find decision-making pathways of medical experts.

References

- [1] Buetow SA, Sibbald B, Cantrill JA, Halliwell S. Prevalence of potentially inappropriate long term prescribing in general practice in the United Kingdom, 1980-95: systematic literature review. *BMJ*. 1996;313:1371-4.
- [2] Coste J, Sene B, Milstein C, Bouee S, Venot A. Indicators for the automated analysis of drug prescribing quality. *Methods Inf Med*. 1998;37:38-44.
- [3] Coste J, Venot A. An epidemiologic approach to drug prescribing quality assessment: a study in primary care practice in France. *Med Care*. 1999;37:1294-307.
- [4] Muijers PE, Janknegt R, Sijbrandij J, Grol PR, Knottnerus JA. Prescribing indicators. development and validation of guideline-based prescribing indicators as an instrument to measure the validation in prescribing behaviour of general practitioners. *Eur J Clin Pharmacol*. 2004;60:739-46.
- [5] Asch SM, Kerr EA, Lapueta P, Law A, McGlynn EA. A new approach for measuring quality of care for women with hypertension. *Arch Intern Med*. 2001;161:1329-1335.
- [6] Balas EA. Trends in hypertensive drug therapy by US office-based physicians. *Hypertension*. 2001;37:E12.
- [7] Weiss R, Buckley K, Clifford T. Changing patterns of initial drug therapy for the treatment of hypertension in a medicaid population, 1997-2000. *Clin Ther*. 2002;24:1451-1462.
- [8] Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, Rubin HR. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA*. 1999;282:1458-65.
- [9] Lomas J, Anderson GM, Dominick-Pierre K, Vayda E, Enkin MW, Hannah WJ. Do practice guidelines guide practice? The effect of a consensus statement on the practice of physicians. *N Engl J Med*. 1989; 321:1306-1311.
- [10] Milchak JL, Carter BL, James PA, Ardery G. Measuring adherence to practice guidelines for the management of hypertension: an evaluation of the literature. *Hypertension*. 2004;44: 602-608.
- [11] Aminzadeh F. Adherence to recommendations of community-based comprehensive geriatric assessment programmes. *Age Ageing*. 2000;29:401-7.
- [12] Stratégie de prise en charge du patient diabétique de type 2 à l'exclusion de la prise en charge des complications. Paris : Agence Nationale d' Accréditation et d' Evaluation en Santé; 2000. Avail. from: <http://www.anaes.fr/anaes/anaesparametrage.nsf/Page?ReadForm&Section=/anaes/SiteWeb.nsf/wRubriquesID/APEH-3YTFUH?OpenDocument&Default=y&>
- [13] WHO Collaborating Centre for Drug Statistics Methodology, ATC Classification index with DDDs. 2005.Oslo 2004.
- [14] Wolfe RM, Sharp LK, Wang RM. Family physicians' opinions and attitudes to three clinical practice guidelines. *J Am Board Fam Pract*. 2004;17:150-7.
- [15] Mosca L, Linfante AH, Benjamin EJ, Berra K, Hayes SN, Walsh BW, et al. National study of physician awareness and adherence to cardiovascular disease prevention guidelines. *Circulation*. 2005;111:499-510.
- [16] Milchak JL, Carter BL, Ardery G, Black HR, Bakris GL, Jones DW, et al. Development of explicit criteria to measure adherence to hypertension guidelines. *J Hum Hypertens*. 2006;16.

Computerised Order Entry Systems: Sustained Impact on Laboratory Efficiency and Mortality Rates?

Johanna I WESTBROOK^{a,1}, Andrew GEORGIU^a, Marilyn I ROB^a

^a *Health Informatics Research & Evaluation Unit, Faculty of Health Sciences, The University of Sydney, 1825, Australia*

Abstract. Few studies have attempted to measure the effectiveness of computerised test-order entry systems to reduce test turnaround time and the extent to which improvements are sustained or continue over time. Further, a recent study has raised the possibility that such systems, which require significant work practice change, may be associated with an increase in mortality rates. Our study answered two questions in relation to the introduction of a computerised pathology order entry system in a major Australian teaching hospital: i) are initial improvements in turnaround times achieved in the first 12 months of system use sustained beyond this time; and ii) did mortality rates change following the introduction of the order entry system? We found significant improvements in turnaround times 12 and 24 months following system implementation and no change in average number of tests per patient. The mortality rate significantly increased in the year following system introduction but returned to the pre-system rate in the second year of system use. Review of the excess deaths demonstrated that these were most likely attributable to a coincidental influenza outbreak and not to system introduction. The computerised order entry systems produced sustained and continuing improvements in health care delivery efficiency over a two year period. Associations between increased mortality rates and system introduction should be investigated carefully to ascertain any likely association.

Keywords. Computer order entry, Evaluation studies, Mortality, Laboratories, Pathology, Turnaround Time, EMR

Introduction

Turnaround time (TAT) is the speed with which a laboratory can process a specimen and provide a result. It is a key performance indicator for laboratories and a potential indicator of quality of care[1]. Reduced TATs are associated with improved clinician satisfaction and shorten lengths of stay in emergency departments[2]. In time- critical care settings the availability of information to support timely diagnostic and management decisions is associated with better outcomes for patients[3].

¹ Johanna Westbrook, Health Informatics Research & Evaluation Unit, Faculty of Health Sciences, The University of Sydney, Cumberland Campus, East Street PO Box 170 Lidcombe NSW 1825 Australia. E-mail: J.Westbrook@usyd.edu.au URL address: www.fhs.usyd.edu.au/hireu

Computerised order entry systems allow clinicians to directly enter orders into a computer and eliminating administrative steps and thus hold promise in significantly reducing test TATs. Few studies have measured TATs before and after system introduction. Westbrook et al[4] demonstrated a 21% reduction in average TAT in the year following system implementation in an Australian hospital and Guss et al[5] showed a similar 20% reduction in TAT six months post-implementation in a US hospital. We have identified no studies which have measured the extent to which initial gains in TAT are sustained beyond the first 12 months of system implementation.

It is now well recognised that major clinical system implementations require significant work practice change within and beyond the laboratories[6-8]. The consequences of these may produce negative effects for care delivery[9] and patient outcomes[10]. The most controversial of these was a reported doubling of the mortality rate from 2.8% before, to 6.6% in the five months following, the introduction of a computerised order entry system in a paediatric hospital in the United States[10]. While there has been considerable debate[11] regarding the contributing roles of the system and the associated work practice changes to the increased mortality, the study drew attention to the possibility that system implementation could result in increased mortality. There is growing acceptance of the need to adopt a multi-method approach to the evaluation of computerised clinical systems[12] in order to capture the multi-dimensional changes which can be produced within complex health care settings. Mortality rates are a further indicator to add to such assessments.

Having published a number of quantitative and qualitative studies[4, 6-8] evaluating the impact of pathology order entry systems during the implementation and early post-implementation periods, our aim in this study was to answer two central questions: i) are initial gains in turnaround times achieved in the first 12 months of system use[4] sustained beyond this time; and ii) did mortality rates change following the introduction of the order entry system?

1. Material and Methods

1.1. Research Setting & Study Design

The study was undertaken at a 650-bed teaching hospital. In November 2003 the Cerner Millennium PowerChart (version 7.8) system was implemented allowing clinicians to electronically order, verify and review pathology orders. The system linked to an existing results-reporting function & hospital patient information systems.

Data relating to discrete test assays ordered during July to August 2003 formed the 'before' database. Data from July to August 2004 formed the first followup period (post 1) and data from July to August 2005 made up the second followup period (post 2). TAT was defined as the time from receipt of a specimen in a laboratory to availability of a result. This definition is most useful in assessing order-entry systems, as the alternate - time from order until result - is confounded by the time taken to obtain and deliver the specimen. This activity is not related to the system but to other aspects of service delivery. Further, within the paper-based ordering system, time of order is inconsistently recorded, reducing accurate pre and post comparisons. Number and details of deaths and average number of tests per patient in each period were abstracted from the order entry and hospital systems. A medical record audit of a sub-group of these patients was undertaken to identify factors contributing to their deaths.

1.2. Analysis

The TAT distribution was extremely skewed. While almost half of the test results were available within 30 minutes and around 80% within one hour, almost 2% took more than three hours and many of these over six hours. To normalize the distribution and obtain more accurate means, a log transformation was undertaken. To test differences in mean times between years, regression analysis was carried out using logarithm scores. Results plus derived 95% CIs were then converted back to natural numbers.

2. Results

2.1. TATs and Mean Number of Tests Before, 12 and 24 Months Post System

Both post-implementation TATs were significantly lower than in the pre-period (Table 1). TATs continued to decline in the second year post-implementation with an average reduction of 12.6% overall (Tables 2 and 3). The average number of tests per patient did not significantly change in the post 1 (mean=109.1) or post 2 periods (115.5; $t=0.577$, $df=2171$, $p=0.5641$).

Table 1 Regression results - pathology turnaround times: 2003 compared with 2004 & 2005

	Period	Param. estimate	Standard Error	t-value	p
All tests	2003	3.56534	0.00350	1019.7	<0.0001
	2004	-0.20603	0.00477	-43.2	<0.0001
	2005	-0.34102	0.00462	-73.8	<0.0001
	R ² =0.0157 N=342625				
Business hours (8am-5pm)	2003	3.82078	0.00384	994.2	<0.0001
	2004	-0.21153	0.00537	-39.4	<0.0001
	2005	-0.32503	0.00523	-62.1	<0.0001
	R ² =0.0172 N=224135				
Non-business hours (5.01pm-7.59am)	2003	2.93529	0.00635	462.0	<0.0001
	2004	-0.02112	0.00826	-2.6	0.0106
	2005	-0.16001	0.00797	-20.1	<0.0001
	R ² =0.0047 N=118490				

2.2. Mortality Pre and Post System Implementation

Comparison of the mortality rate before and in the two followup periods showed a significant increase in deaths in 2004 relative to 2003 (Chi Sq=4.54, $df=1$, $p=0.033$). The rate then declined in 2005 to a rate similar to the pre-system implementation mortality rate. Thus when examined across the three year periods there was no significant increase in mortality rates (Table 4).

Table 2 Regression results - pathology turnaround times: 2004 compared with 2005

	Period	Parameter estimate	Standard Error	t-value	p
All tests	2004	3.35932	0.00325	1033.9	<0.0001
	2005	-0.13500	0.00444	-30.4	<0.0001
	R ² =0.038 N=244773				
Business hours	2004	3.60924	0.00381	947.0	<0.0001
	2005	-0.11350	0.00524	-21.6	<0.0001
	R ² =0.0030 N=154511				
Non-business hours	2004	2.91417	0.00525	554.9	0.0106
	2005	-0.13889	0.00710	-19.3	<0.0001
	R ² =0.0042 N=90263				

Table 3 Mean turnaround times and 95% Confidence Intervals, 2003, 2004 & 2005

	Period	No. of tests	Mean in minutes (95% C.I.)
All tests	2003	97851	35.35 (35.11,35.59)
	2004	113752	28.77 (28.59,28.95)
	2005	131022	25.14 (24.99,25.29)
Business hours	2003	69624	45.64 (45.31,45.97)
	2004	72850	36.94 (36.67,37.21)
	2005	81661	32.97 (32.74,33.21)
Non-business hours	2003	28227	18.83 (18.59,19.07)
	2004	40902	18.43 (18.25,18.62)
	2005	49361	16.04 (15.89,16.20)

While not statistically significant the increase in deaths in the first year of system use was of clinical significance and deemed worthy of investigation. In 2004 there were 37 excess deaths compared to the year prior. An examination of patient demographics showed that 69.4% of patients who had died were over 80 years, with the exception of one. Further analyses showed all excess deaths in 2004 had occurred among those over 80 years (Table 5).

Examination of patients' Diagnosis Related Groups (DRG) revealed the excess deaths were limited to those in the DRG 'E' group, respiratory conditions, and again only among patients aged 80 years or older (Table 6).

Table 4 Number of deaths before and in the two followup periods

Period	No. of patients	No. (%) who died	Chi Sq	df	p
Jul/Aug 2003	9763	149 (1.53)	5.5603	2	0.0620
Jul/Aug 2004	9394	181 (1.93)			
Jul/Aug 2005	10100	159 (1.57)			

Table 5 Changes in mortality rates among those under, and ≥ 80 years 2003-2005

Age of patient	Period	No. of patients	No. (%) who died	Chi sq	df	p
Under 80 years	Jul/Aug 2003	9163	114 (1.24)	1.2724	2	0.5293
	Jul/Aug 2004	8723	113 (1.30)			
	Jul/Aug 2005	9394	105 (1.12)			
80 years or older	Jul/Aug 2003	600	35 (5.83)	8.1429	2	0.0171
	Jul/Aug 2004	671	68 (10.13)			
	Jul/Aug 2005	706	54 (7.65)			

Table 6 Changes in mortality rates for patients under, and ≥ 80 years with respiratory conditions 2003-2005

Age of patient	Period	No. of patients	No. (%) who died	Chi sq	df	p
Under 80 years	Jul/Aug 2003	516	21(4.07)	0.8160	2	0.6650
	Jul/Aug 2004	395	20 (5.06)			
	Jul/Aug 2005	414	16 (3.86)			
80 years or older	Jul/Aug 2003	89	7 (7.87)	12.0037	2	0.0025
	Jul/Aug 2004	94	26 (27.66)			
	Jul/Aug 2005	79	15 (18.99)			

It was hypothesised that the increase in deaths may have been associated with an influenza outbreak in 2004. We reviewed cases of Pneumococcal disease for the State and found cases increased from 808 in 2003 to 917 in 2004 and then dropped to 660 in 2005, supporting this hypothesis. An audit of patients' medical records failed to reveal any system-related factors associated with deaths. Most patients had significant comorbidities, came from nursing homes or hostels and had very poor prognoses.

3. Discussion

The use of the computerised order-entry system led to significant reductions in test turnaround times in the 12 months following implementation and these improvements were sustained and improved upon in the following 12 months. Further research is required to demonstrate whether faster results enhance patient outcomes. Although the initial comparison of crude mortality rates suggested a link between the introduction of the new system and the (non-significant) rise in death rate, this rise was limited to a

very specific group of patients whose deaths could be attributed to the influenza outbreak which occurred coincidentally after system implementation.

4. Conclusion

Computerised order entry systems can deliver sustained improvements to health care efficiency. Mortality rates should continue to be examined as one evaluation indicator but investigated carefully. An initial association between system implementation and mortality rate does not equate to causation.

5. Acknowledgements

The research was funded by an ARC Linkage Grant in partnership with NSW Health. JW is supported by a NHMRC Fellowship. The authors thank A Ampt and C Roberts.

References

1. Manor, P., *Turnaround times in the laboratory: A review of the literature*. Clin Lab Sci, 1999. **12**(2): p. 85-89.
2. Lee-Lewandrowski, E., et al., *Implementation of a point-of-care satellite laboratory in the emergency department of an academic medical center. Impact on test turnaround time and patient emergency department length of stay*. Arch Path Lab Med, 2003. **127**(4): p. 456-460.
3. Richardson, D., *The access-block effect: relationship between delay to reaching an inpatient bed and inpatient length of stay*. Med J Aus, 2002. **177**: p. 492-495.
4. Westbrook, J., et al., *Computerised pathology test order entry reduced laboratory turnaround times and influences tests ordered by hospital clinicians: a controlled before and after study*. J Clin Pathol, 2006. **59**: p. 533-536.
5. Guss, D., T. Chan, and J. Killeen, *The impact of a pneumatic tube and computerized physician order management on laboratory turnaround times*. Ann Emerg Med, 2008. **51**(2):p. 181-185
6. Georgiou, A., et al., *Multiple perspectives on the impact of electronic ordering on hospital organisational and communication processes*. Hlth Inform Mngt, 2005. **34**(4):p. 130-135.
7. Georgiou, A., et al., *When requests become orders - a formative investigation into the impact of computerised physician order entry systems on a pathology service*. Int J Med Inform, 2007. **76**: p. 583-591.
8. Georgiou, A., et al., *The impact of computerised physician order entry systems on pathology services: a systematic review*. Int J Med Inform, 2007. **76**(7): p. 514-529.
9. Ash, J., M. Berg, and E. Coiera, *Some unintended consequences of information technology in health care: The nature of patient care information systems related errors*. J Am Med Inform Ass, 2004. **11**(2): p. 104-112.
10. Han, Y., et al., *Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system*. Ped, 2005. **116**: p. 1506-1512.
11. Ammenwerth, E., et al., *Impact of CPOE on mortality rates - contradictory findings, important messages*. Meth Infom Med, 2006. **45**: p. 586-594.
12. Westbrook, J., et al., *Multi-method evaluation of information and communication technologies in health in the context of wicked problems and socio-technical theory*. J Am Med Inform Ass, 2007. **14**(6): p. 746-755.

5. Health Information Systems Including EHR

This page intentionally left blank

Managing Care Pathways combining SNOMED CT, Archetypes and an Electronic Guideline System

Knut BERNSTEIN^a, Ulrich ANDERSEN^b

^a *MEDIQ – Medical Informatics and Quality Management, Denmark.*

^b *International Health Terminology Standards Development Organisation, Denmark*

Abstract: Today electronic clinical guideline systems exist, but they are not well integrated with electronic health records. This paper thus proposes that the patient's "position" in the pathway during the patient journey should be made visible to all involved healthcare parties and the patient. This requires that the generic knowledge, which is represented in the guidelines, is combined with the patient specific information – and then made accessible for all relevant parties. In addition to the decision support provided by the guideline system documentation support can be provided by templates based on archetypes. This paper provides a proposal for how the guideline system and the EHR can be integrated by the use of archetypes and SNOMED CT. SNOMED CT provides the common reference terminology and the semantic links between the systems. The proposal also includes the use of a National Patient Index for storing data about the patient's position in the pathway and for sharing this information by all involved parties.

Keywords: Evidence based guidelines, terminology, knowledge management

1. Introduction

Important challenges in modern healthcare are to provide continuity of care and evidence based care. This implies that care should be based on relevant and updated knowledge that is provided in a timely fashion and delivered coherent across professional and organisational borders. The use of clinical guidelines that describe the sound care pathways and are based on clinical evidence is one of the major instruments to achieve this.

This paper proposes that the patient's "position" in the pathway during the patient journey should be made visible to all involved healthcare parties and the patient. This requires that the generic knowledge, which is represented in the guidelines, should be combined with the patient specific information – and then made accessible to all relevant parties. The integration with the clinical IT systems – i.e. the electronic health records (EHR) – also includes the semantic interoperability, which is a substantial challenge. However, the semantic interoperability is needed to ensure safe exchange of data between the guideline system, the documentation templates and the EHR.

This paper provides a proposal for how the guideline system and the EHR can be integrated by using SNOMED CT and archetypes.

2. Method

The aim of this paper is to present a proposal for improvements of the semantic interoperability between electronic guideline and EHR systems. The proposal is based on an analysis of the healthcare challenges, providing continuity of care and evidence based care.

The idea is to use existing logical and technical infrastructure as a foundation. Key components of such an infrastructure will therefore be described – including a common terminology, an electronic guideline system and a national patient index.

Based on these infrastructure components, a generic solution is described. This contains an overall architecture, showing how the components interwork. The hypothesis is that this solution will support the resolution of the healthcare challenges.

3. Result

3.1. Problem and significance

Modern healthcare is specialised and many healthcare providers are involved in the patient journey. Coherent provision of care across professions and organisational and geographical borders have thus become a challenge. At the same time political goals on containing the efficiency and quality should be observed. The use of evidence-based guidelines is considered a way to achieve these goals. [1]

It has been demonstrated that clinical guidelines provided as real time decision support systems in daily clinical workflow will improve patient care significantly [2, 3] and are effective instruments to decrease undesired practice variability [4]. The success of clinical decision-support systems requires that they are seamlessly integrated with clinical workflow [5].

3.2. Proposed solution

An overview of the proposed solution is shown in Figure 1. The idea is that a healthcare professional (HCP) “place” the patient at the relevant position in the pathway and the position in the pathway is changed when the patient is going through the various steps. The information about the current position in relation to the whole patient journey is shared by all involved HCP and the patient. Information about the patient journey and the correlation to the recommended pathway is collected in a National Patient Index (NPI). This way compliance with recommendations and coherence with minimum waiting times (i.e. in cancer pathways) can be monitored by all involved parties. The generic knowledge from the guideline system is used to decide the right pathway for a specific patient. The documentation templates (i.e. archetypes) are used to support registration of the correct data.

The specific patient information – including the reference to the step in the pathway – is stored in the EHR. This is reported to the NPI, which can match the

specific patient journey with the recommended pathway. The terminology system SNOMED CT is used to provide exact links between the node in the guideline system, the documentation template and the information in the EHR, figure 1.

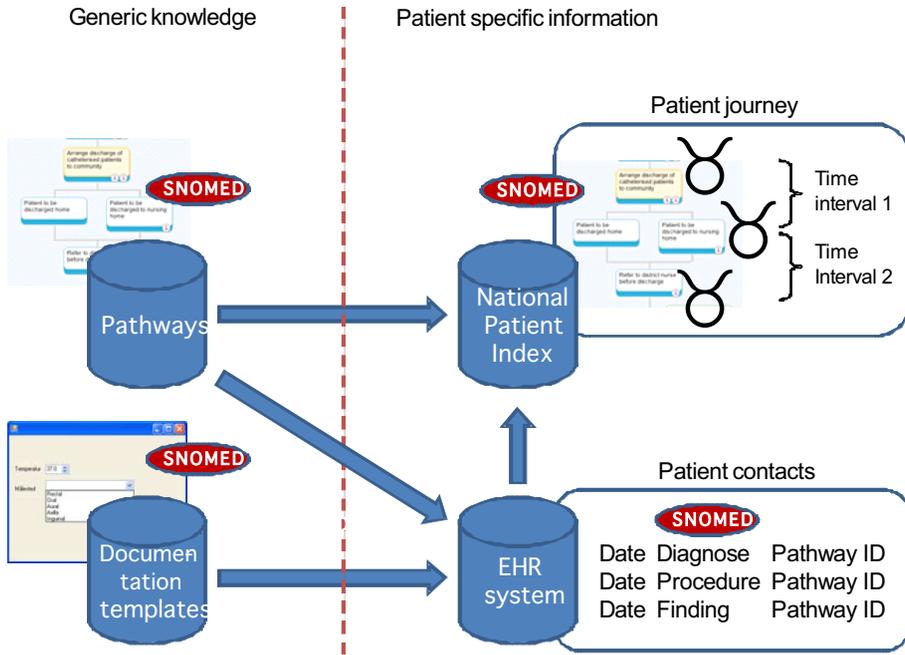


Figure 1. How to manage care pathways by combining SNOMED CT, archetypes and an electronic guideline system

3.2.1. Basic components

The basic components of the infrastructure are described below. In many countries these components already exist or are being implemented. Some adjustments are needed in the EHR systems and the NPI to handle the data from the guideline system – however, only very few data types are needed in the basic configuration. If documentation templates based on archetypes are chosen, non-archetype systems probably need larger adjustments to make them fit the overall solution.

National Patient Index

A National Patient Index (NPI) is in this context a common database with core patient data, i.e. a unique patient identifier, all contacts (consultations, encounters) with dates, diagnoses, procedures and an identification of the relevant node in the pathway. In addition, the NPI contains links to more detailed patient information i.e. findings, results, images etc. What information is stored centrally and what information can be reached by links to external sources depends on national choices and infrastructure.

Electronic guideline system, i.e. Map of Medicine

Map of Medicine [6] is an implementation of a guideline system and it is used as an example of an application in this paper. It is being used by the NHS in UK and it is

evaluated in Denmark and other countries. Map of Medicine is a web based system, providing specialist level knowledge to healthcare professionals. The user interface is intentionally kept simple, displaying the pathways as boxes linked together like a process flow diagram. There is only one type of box for a node, colour coded for use in primary or secondary care setting. For each node additional information can be displayed, i.e. detailed guidance, checklists etc. The evidence base for each node and guideline is directly accessible, as are links to medical databases, Google etc. All pathways and nodes can be customised at a national, regional, local and personal level.

SNOMED CT®

SNOMED Clinical Terms (SNOMED CT) is a structured healthcare terminology, consisting of more than 300.000 unique medical concepts and more than 1.000.000 terms. [7] SNOMED CT offers a terminological foundation for EHR and other health IT systems. The international release of the terminology is managed by The International Health Terminology Standards Development Organisation (IHTSDO) [8]. IHTSDO was founded by nine countries and SNOMED CT is currently translated into several languages, [9] becoming the global standard of healthcare terminologies.

Archetypes

The archetypes discussed in this paper are based on the openEHR reference model [10]. This is equivalent to the CEN pre-standard 13606 (EHRcom), which is currently under revision [11]. OpenEHR describes archetypes as "... reusable, structured models of clinical information concepts that appear in EHRs". Archetypes are described in a formal language. Archetypes are used to create (small) clinical models of for example observations like blood pressure or a specific laboratory result, information related to the evaluation of a clinical problem, related to prescribing of medications etc.

The templates discussed are based on the openEHR concept where archetypes can be adjusted to local needs and connected to entry forms (documentation templates), (to) terminologies, language and other details relevant to the particular local use. OpenEHR templates can be used to design documentation templates for example hospital admission, where an archetype for vital signs can be combined with an archetype for patient history and specific entry fields i.e. for diagnoses that can be linked to a relevant SNOMED CT subset [12].

3.2.2. Linking the components

Connecting the components with semantically rich links in a real-time, dynamic fashion is a challenge. SNOMED CT is proposed as the mechanism to achieve this.

Using SNOMED CT in EHR systems for diagnoses, procedures and findings enables the SNOMED CT code to work as a parameter to perform a precise look-up in the guideline system. Furthermore, a specific node in the pathway may be connected to a specific contact in the EHR or even to a specific diagnose, procedure or finding. By combining the contact data with a unique ID of the node in the pathway, it is possible to identify the patient current position and change in positions in the guideline during the patient's journey.

SNOMED CT could also be the link to the documentation templates via the archetypes. Each archetype is intended for use in a specific clinical situation and a procedure selected in the EHR or in the guideline system could trigger a documentation template in the EHR system. The data recorded via the template will be stored in the

EHR according to the data structure of the EHR system. Furthermore, data entry fields in the documentation template may be linked to SNOMED CT via the archetype/template definition.

As mentioned, the patient's current position in the pathway can be identified. By adding this "position information" to the patient data reported to the NPI, it is possible to share information concerning any progress in the patient's journey.

3.2.3. Resulting functionality - scenarios

The generic functionality arising from the linked components has been indicated above. In the tables below example scenarios are listed from the healthcare professional, the authority/administrator and from the patient/relatives viewpoints.

Table 1. Examples of scenarios related to the healthcare professional

Decision support: The physician is suspecting diabetes and decides to initiate a screening for diabetes. The procedure "diabetes screening" is selected in the EHR and this term has the relevant SNOMED CT code attached. When Map of Medicine is launched (i.e. with a mouse right-click), the code is used to search the Map of Medicine. The guideline "diabetes suspected" is presented to the physician.
Documentation support: One of the first steps in the guideline is "taking the patient history". The guideline contains an extensive list of information to be collected. The node has a SNOMED CT code attribute, which is used to identify the archetype/documentation template to be used by the physician in this step of the guideline.
Process support: After the history taking and physical examination, some laboratory investigations are ordered. The physician registers that the patient now has reached this node in the guideline and is waiting for results. This status of the patient is reported to the National Patient Index together with the key information from the patient contact, i.e. the diagnosis and the procedure. This information is now accessible (i.e. via a health portal) for all involved healthcare actors and the patient.

Table 2. Examples of scenarios related to the authorities

Core data on a patient's treatment: Based on archetypes semantically well defined data can be reported to national databases. Since the documentation based on archetypes/templates gives an explicit specification of the information context, it is also safer to use data for secondary purposes, i.e. quality databases.
Precise positioning in pathway: By connecting the patient contact to the Map of Medicine, the various steps in the patient journey can be identified with high precision. This gives an opportunity for two kinds of quality assessments. Firstly, statistics can be made in the time interval between nodes in the guideline. This is important if minimum waiting times exist, i.e. for cancer treatment. Secondly, systematic deviation from the guideline can be identified and explanations can be explored.

Table 3. Examples of scenarios related to the patient

Knowledge of pathways: The patient has an authoritative information source on clinical relevant pathways related to the diagnosis or treatment. This can be used for seeking further information from the physician.
Follow up on own patient journey: The patient journey is transparent(.) and the patient can follow the steps in the pathway. This can help to empower the patient to take action if i.e. waiting times are too long

4. Discussion

The advantages with the proposed solution are illustrated by the scenarios in table 1-3. The transparency of the patient journey supports the continuity of care.

The use of SNOMED CT supports the accurate linking between general guideline knowledge and patient specific information. The use of templates/archetypes supports

documentation of the information relevant to the specific clinical situation, i.e. the node in the guideline. Access to core information via the NPI enables the follow-up on political and clinical goals and supports patient empowerment.

There are also drawbacks and barriers related to the proposed solution. Significant resources are needed for setting up the clinical infrastructure, i.e. development, maintenance, possibly translation of the guidelines, archetypes and the terminology system. Furthermore, resources are needed for setting up the technical infrastructure, i.e. the interfaces, standards, systems and services. Also a secure environment regarding the patient specific information has to be established.

The system depends on the HCP registering the link between the pathway and the patient. This implies new work routines and maybe additional effort. It is therefore imperative that the systems are providing perceived advantages for the HCP, i.e. regarding efficiency and quality.

5. Conclusion

By using electronic guideline system integrated with EHR there is a potential for providing better decision support, documentation support, process support to the HCP, quality improvement and improved service to the patient. The proposed solution to use SNOMED CT, archetypes and a NPI requires significant clinical and technical infrastructure.

However, several of these components are already available. Thus, a stepwise implementation is possible. For instance, in Denmark where an NPI exists and the SNOMED CT translation is almost finalised. Furthermore, the Map of Medicine and an archetype methodology is evaluated to prepare for a possible implementation.

6. References

- [1] Stein M. The Map of Medicine – an Innovative Knowledge Management Tool. AMIA 2006 Symposium proceedings, page 1196
- [2] Kawamoto K et al. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ*. 2005 Apr 2;330(7494):740-1).
- [3] Quaglini S, Ciccarese P. Models for guideline representation. *Neurol Sci*. 2006 Jun;27 Suppl 3:S240-4
- [4] Lenz R, Blaser R et al. IT support for clinical pathways--lessons learned. *Stud Health Technol Inform*. 2006;124:645-50
- [5] Tu SW, Musen MA et al. Modeling guidelines for integration into clinical workflow. *Medinfo*. 2004;11(Pt 1):174-8
- [6] <http://www.mapofmedicine.com/>
- [7] SNOMED Clinical Terms User Guide. July 2006 Release. College of American Pathologists
- [8] www.ihtsdo.org
- [9] Ulrich Andersen, Janni Lerche, Palle Gerry Petersen, Knut Bernstein. Adapting SNOMED CT® for use in Denmark – the tools and the process of concept based translation. *Medinfo*. 2007;12:4047-9
- [10] Introducing openEHR. openEHR, release 1.0. 2005
- [11] CEN /TC251 prEN 13606-2. Health informatics - Electronic health record communication - Part 2.
- [12] Bernstein K, Darmer MR. Representing Clinical Knowledge as Archetypes. *Medinfo*. 2007;12:3078-80

Analysis and Evaluation of EHR Approaches

Bernd G.M.E. BLOBEL¹ and Peter PHAROW

eHealth Competence Center, Regensburg University Hospital, Regensburg, Germany

Abstract. EHR systems are core applications in any eHealth/pHealth environment and represent basic services for health telematics platforms. Many projects are performed at the level of Standards Developing Organizations or national programs, respectively, for defining EHR architectures as well as related design, implementation, and deployment processes. Claiming to meet the challenge for semantic interoperability and offering the right pathway, the resulting documents and specifications are sometimes controversial and even inconsistent. Based on a long tradition in the EHR domain, on the collective experience of academic groups such as the EFMI EHR Working Group, and on an active involvement at CEN, ISO, HL7 and several national projects around the globe, an analysis and evaluation study has been performed using the Generic Component Model reference architecture. Strengths and weaknesses of the different approaches as well as migration pathways for re-using and harmonizing the available materials are offered.

Keywords. EHR, Architecture, Models, Semantic Interoperability, Standards, National Projects.

Introduction

The requirements for safe and high quality care as well as efficiency and productivity of health systems under the well-known constraining conditions are expected to be realized by increasingly distributed and specialized health services that become strongly oriented on the actual personal health status and the needs of the subject of care. Those health services are provided independent of time, localization, and distribution of resources in a highly communicative and collaborative way called eHealth. eHealth or personal health must be supported by basic technology paradigms of mobile computing for ubiquitous communications, of pervasive computing for comprehensive, and of pervasive care as well as autonomic computing for adaptive personalized system design enabling ubiquitous care altogether. Intentional communication and cooperation needs shared information for deriving collaborative actions which have to be observed for Quality Assurance (QA) purposes starting the information cycle again and again. Principles and consequences considering the entire information cycle for meeting the semantic interoperability challenge are extendedly discussed in [1]. According to the

¹ Corresponding Author: Bernd Blobel, PhD, Associate Professor; eHealth Competence Center, Regensburg University Hospital; Franz-Josef-Strauss-Allee 11, D-93053 Regensburg, Germany; Phone: +49-941-944-6769. Fax: +49-941-944-6766. Email: bernd.blobel@klinik.uni-regensburg.de; <http://www.ehealth-cc.de>

objectives and requirements of the actors involved in communication and cooperation, different levels for interoperability shall be provided thereby extending the interoperability chain up to the subject of care as demonstrated in the personal care paradigm's context [2]. Depending on the business case requiring interactions between systems, the following interoperability levels can be distinguished: technical interoperability (e.g. technical Plug&Play, signal & protocol compatibility); structural interoperability (e.g. simple EDI, envelopes); syntactic interoperability (e.g. messages, clinical documents, agreed vocabulary); semantic interoperability (e.g. advanced messaging, common information model and terminology); organizational/service interoperability (e.g. the common business process). The information about status and processes directly or indirectly related to the subject of care has to be documented in a computer-readable format, and stored in repositories - the Electronic Health Record (EHR). Such EHRs have to be implemented in EHR systems which are sets of components for realizing the mechanisms for creating, using, storing, and retrieving an EHR. EHRs and EHR systems for eHealth have to be based on an EHR architecture defining a model of generic properties required for every EHR to be communicable, comprehensive, useful, effective, and ethically-legally binding bearring its integrity independent of platforms and systems as well as national specialties over the time [3], [4].

1. Characteristics of Semantically Interoperable EHR Architectures

For providing advanced and sustainable communication and cooperation, architectures for sustainable Health Information Systems (HIS) such as EHR systems have to be open, scalable, flexible, portable, distributed, standard-conform, semantically interoperable, service-oriented, user-accepted, applicable to any media, and lawful. Therefore, the following architectural paradigms have to be met: distribution; component-orientation; a model-driven and service-oriented design considering concepts, context, and knowledge; comprehensive business modeling; separation of platform-independent and platform-specific modeling (thus separating logical from technological view); agreed reference terminologies and ontologies; unified development process, and advanced security and privacy services embedded in the architecture. The aforementioned architectural paradigms are reflected in the Generic Component Model (GCM) which provides a multi-model approach to any system architecture [5]. Developed at the Magdeburg Medical Informatics Department in the early nineties, it can be used as reference architecture for analyzing, designing, and implementing EHR architectures characterized by their components, functionalities, and relationships but also for developing migration strategies [5], [6].

The respective reference architecture design is characterized by three dimensions (see Figure 1):

- the composition and decomposition of its components reflecting the complex functionality; from an enterprise viewpoint (explained in the next statement), it's the business process (workflow);
- the unified development process referring to the viewpoint of the ISO 10746 Information Technology – Open Distributed Processing – Reference Model [7] starting with the extended business modeling of the enterprise view, the information view as its informational reflection, and the computational view as the reasonable functional aggregation of the components all three reflecting the platform-independent modeling and continuing with the platform-specific

models of the engineering view and the technology view, all five views summarized as development process (e.g. the Rational Unified Process – RUP);

- the isolation and integration of the domain touched by the EHR (medical, administrative/organizational, legal, etc.).

EHR systems as well as standards and projects defining them have to be assessed in reference to the GCM reference architecture. So, the usability of the analyzed approach, the gaps, and the capability for migration can be evaluated; the respective migration paths can be derived.

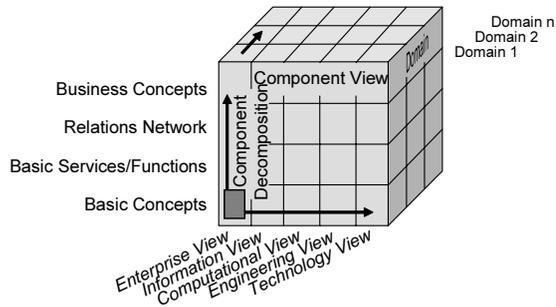


Figure 1. The Generic Component Model

2. EHR-Related Standards

Currently, three different streams for specification and implementation of advanced EHR architectures exist that have their roots in legacy systems, traditional imaginations, and methodologies: data approach (data representation), concept approach (concept and knowledge representation), and process/services approach (business process and service representation).

Because of their rational roots, all approaches have their -at least temporary- right to exist. All approaches undergo further development offering a convergence. As the GCM considers all aforementioned aspects, the distance of evaluated approaches to the GCM as well as the reflection of the presented principles allow for the evaluation of developed or emerging solutions as well as for description of missing characteristics.

Regarding EHR approaches as standards, specifications, or national projects, the following ones have to be investigated as the most important ones: the HL7 standards suite [8] with the HL7 RIM, HL7 v3, HL7 HDF, CDA, etc. most of them established, or under establishment, in ISO TC 215; ISO/EN 13606 “Health informatics – EHR communications” [9]; the GEHR [10] and openEHR Foundations [11] projects; IHE approaches [12] but also service-oriented approaches such as CORBA components [13]; Web services, etc. The analysis only focuses on the architectural approaches ignoring national or project-specific specialties on implementation details, legal aspects, or priorities followed in some countries and their related projects.

3. Results

Using the Generic Component Model, the different standards and projects influencing the global EHR development have systematically been evaluated. The results are summarized in Table 1. While in the study [14] all characteristics have been analyzed in detail, the table only reflects coarse grained quality parameters not considering either how the different aspects elaborated have been harmonized within the standards (quality and consistency). Due to the complexity of the ISO organization, standards might

be referenced responding to some of the GCM requirements not inter-relating with EHR specifications, however, while the HL7 standard set provides such inter-relationships as well as formal collaboration and joint projects solving those inter-relations between SDOs. In that context, the unified process, separations of viewpoints, and the completeness of referencing ISO 10746-2 [7] have to be mentioned which are met in HL7 but not in CEN/ISO or openEHR, even if standards such as ISO EN 12967 [15] are sometimes miss-referenced there.

Table 1. Comparison of available EHR approaches (availability: P-partial, Y-yes, N-no, F-future)

GCM Characteristics	HL7 Standards	CCR	EN/ISO 13606	openEHR	IHE XDS	DICOM SR
Development Process						
Unified process	Y	N	N	N	N	N
Business modeling	Y	N	N	N	P	N
Service orientation	P	N	N	N	P	N
View separation	P	N	N	N	N	N
Completion of ISO 10746-2	Y	N	N	N	N	N
System Architecture						
Reference information model	Y	N	Y	Y	Y	Y
Meta model	P	N	N	N	N	N
Model transformation framework	Y	N	N	N	N	N
Concept representation	Y	P	Y	Y	N	Y
Consistency of components	N	Y	Y		Y	Y
Open concept representation language	Y	Y	N	N	N	N
Composition/decomposition	Y	N	N	N	N	N
Signature/Certificate-enabled	N	Y	N	N	Y	Y
Machine-processable	P	P	P	P	Y	Y
Multi-Domain Suitability						
Domain-independent	N	N	N	N	N	N
Domain separation	N	N	N	N	N	P
Model multiplicity	n	1	2	2	2	2
Ontology driven	N	N	Y	Y	N	N
Vocabulary	Y	N	N	N	N	N
Reference to terminology	Y	Y	Y	Y	Y	Y
Communication security services	N	Y	N	N	Y	Y
Application security services	N	Y	P	N	N	N
Inclusion of medical devices	Y	N	F	N	F	Y
Specialty-related	N	Y	N	N	Y	N
Multimedia-enabled	P	P	N	N	Y	Y
Feasibility						
Visualization support	Y	Y	N	N	N	N
Final specification available	N	Y	N	N	Y	Y
Implemented	Y	Y	N	P	Y	Y
Commercial products available	Y	Y	N	N	Y	Y

4. Discussion and Conclusions

The core application EHR is in the center of considerations for all European (e.g. EC eHealth Action Plan) and national (UK, Denmark, USA, Finland, Australia, Canada) eHealth Programmes. With the move by a paradigm change from organization-centered to process-controlled care and even further towards personalized healthcare, comprehensive communication and cooperation between all participants in the process including semantic interoperability between supporting information systems is inevitable.

Different advanced approaches for future-proof architectures, EHR specifications, and the implementation of semantically interoperable EHR systems (e.g. HL7 Version 3 Standard Set with CDA, CCD, EHR-S Functional Model, EHR Interoperability Model [8], GEHR [10] /openEHR [11], EN/ISO 13606 [9], and CCR [16]) have been demonstrated, discussed, and evaluated using the GCM as reference architecture for sustainable and semantically interoperable health information systems [5].

The HL7 Version 3 methodology in connection with the definition of system requirements by the EHR-S Functional Model and the EHR-S Interoperability Model (it

remains a question to the author why both models have not been brought together) provides the best approach so far to GCM without solving domain-crossing aspects and the connection of non-ICT views to ICT views. Furthermore, the formal business process specification and the dynamic behavioral/functional aspects of the components are still missing. The service-orientation missing could be overcome by current efforts of the SOA SIG in liaison [9]. Additionally, present concept representations have not been adequately integrated, probably due to a missing HL7 ontology. Surmounting the many solution islands, the complex HL7 Standards family could -in collaboration with CEN, openEHR and CORBA- demonstrate some progress.

The second rather comprehensive approach to semantic interoperability is EN/ISO 13606 [9], even if many deficiencies and inconsistencies have yet to be overcome. Contrary to the HL7 Version 3 approach, the problem of semantic composition / decomposition is insufficiently solved there. The same counts for business processes. On the other hand, the project orientates right from its beginning to the architecture paradigm despite of the irritating title, and it goes beyond the HL7 approach in this perspective. GEHR/openEHR [10], [11] has to be evaluated partly analogue to EN/ISO 13606 due to the close connections and the common knowledge representation based on Archetypes. It constraints itself in essence to parts 2 and 3, however.

These two promising approaches suffer from the complexity of the healthcare domain. So, it will take some time before a critical mass of model-based services (meta-models at different levels or Archetypes, respectively, as well as tools for instantiation) will be available for revolutionizing health. Since 2001, different pathways have been followed. While EN/ISO 13606 still focuses on its architectural approach, the openEHR Foundation project focuses on concepts bound to the clinical process leading to different structural components. At the same time, commonalities with HL7 CDA and HL7 Clinical Templates are increasing [8].

Contrary to HL7 Version 3 Standard Set and EN/ISO 13606, CCR [16] provides an immediately applicable record solution without claiming semantic interoperability, however. All ways offered allow for migration using the GCM. A closer cooperation between standards bodies is absolutely helpful and thus required. During the evolution, the user community has to decide which interim solutions are acceptable.

The necessity of meeting all paradigms of the GCM has been emphasized through experience from national projects in different countries the author has been involved in. Having been the internationally leading programme for introducing a national EHR for quite a long time, the Danish approach started with the underlying business processes in an exemplary way. The problem of structural and functional composition / decomposition has been ignored, however. This deficiency stopped the project now resulting in a comprehensive restart. Other projects and standards including the ones discussed here ignore the business processes, at the same time providing reasonable solutions for the other aspects. This has adverse effects as well in case the gaps have not been single-mindedly closed [17].

An important requirement for achieving semantic interoperability has been the establishment of a unified process including the definition of conformance statements as well as the quality assurance for specifications and implementations. Here, projects such as the European Q-REC led by the EuroRec Institute, and the work of the US Certification Commission for Health Information Technology (CCHIT) do push testing, quality labeling, or certification, respectively, of EHR specifications and EHR systems [18], [19].

Currently, not any single specification investigated meets from an insider's perspective the requirements for semantic interoperability at service level – but this is what almost all of them claim to do. The maturity of the approaches is very different whereby history of specifications and originating organizations, the chosen paradigm as well as scope and objectives are of importance. Maturing and harmonizing the present approaches in accordance with the GCM will provide the sustainable solution. Thereby, all architectural paradigms have to be met. Such component-based and service-oriented approach will lead to a layer of infrastructural services supporting creation, deployment, and maintenance of EHR systems. Most of the national strategies promote such development.

The eHealth Competence Center (eHCC) located at the University of Regensburg Medical Center is involved in most of the international standardization activities and national programs related to EHR. While the work in many regions of the globe is ongoing and -more or less- evolved, the German EHR development starting with the bIT4health (better IT for better health care) project that was launched some years ago by the German Federal Ministry for Health is comparably immature. Recently, the eHCC was appointed to specify the German EHR architecture.

Acknowledgement

The authors are indebted to their colleagues from standards bodies and institutions such as ISO, CEN, HL7, the EuroRec Institute as well as related projects like GEHR/openEHR and Q-REC for kind cooperation and support.

References

- [1] Blobel B: Introduction into Advanced eHealth – The Personal Health Challenge. In: Blobel B, Pharow P, Nerlich M (Eds.): eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge – The Global Experts Summit Textbook. Series “Studies in Health Technology and Informatics”, Vol. 134. IOS Press Amsterdam, Berlin, Oxford, Tokyo, Washington, DC, 2008.
- [2] Blobel B, Pharow P, Norgall T. How to Enhance Integrated Care towards the Personal Health Paradigm? In: Kuhn KA, Warren JR, Leong T-Z (Eds.): MEDINFO 2007, pp. 172-176. IOS Press Amsterdam, Berlin, Oxford, Tokyo, Washington, DC, 2007.
- [3] ISO DTR 20514 “Health informatics - Electronic Health Record – Definition, Scope and Context”
- [4] ISO TS 18308 “Health informatics - Requirements for an Electronic Health Record Architecture”
- [5] Blobel B. Advanced EHR Architectures – Promises or Reality. *Meth. Inf Med* 2006; 45: pp 95-101.
- [6] Blobel B: Educational Challenge of Health Information Systems' Interoperability. *Methods Inf Med* 2007; 46: pp. 52-56.
- [7] ISO 10746 “Information Technology – Open Distributed Processing”: <http://www.iso.org>
- [8] Health Level Seven Inc.: <http://www.hl7.org>
- [9] EN/ISO 13606 “Health Informatics – EHR Communications”: <http://www.centc251.org>
- [10] Australian Ministry for Health and Aging: The GEHR Project: <http://www.gehr.org>
- [11] Beale T. A Model Universe for Health Information Standards (2003): <http://www.deepthought.com.au>
- [12] Integrating the Healthcare Enterprise: www.ihe.net, see also www.rsna.org
- [13] Object Management Group, Inc.: <http://www.omg.org>
- [14] DIN Project. Internal Working Document on EHR Requirements. eHCC / DIN Germany.
- [15] ISO CEN 12967 Health Informatics - Service Architecture
- [16] American Society for Testing and Materials: <http://www.astm.org>
- [17] Bernstein K, Bruun-Rasmussen M, Vingtoft S, Andersen SK, Nøhr C. Modelling and Implementing Electronic Health Records in Denmark. *International Journal of Medical Informatics*, 2004
- [18] The Eurorec Institute: <http://www.eurorec.org>
- [19] Certification Commission for Healthcare Information Technology: <http://www.echit.org>

Electronic Disease Surveillance for Sensitive Population Groups – The Diabetics Case Study

Taxiarchis BOTSIS^{a, 1, 2, 3}, Ole HEJLESEN^b, Johan Gustav BELLIKA^{a, b} and Gunnar HARTVIGSEN^{a, b}

^a *Department of Computer Science, University of Tromsø, Norway*

^b *Department of Health Science and Technology, Aalborg University, Denmark*

Abstract. Diabetics are quite susceptible to infectious diseases and can easily spread them under certain circumstances. Their blood glucose levels are increased after infection and this can cause a hyperglycemic crisis. Our study indicates that this increase results in glucosylated hemoglobin elevation, even when a diabetic is monitored closely and his/her blood glucose is under tight control. Thus, it is important to detect infections at the very early stages of disease progression in order to aid the patient. For this purpose, an electronic Disease Surveillance System could be developed to collect and analyze blood glucose data. Generally, we could extend the use of blood glucose data to the implementation of disease surveillance systems for the general population.

Keywords. disease surveillance, early detection, diabetics, blood glucose.

Introduction

Over the centuries many incidents of pandemics have been reported that resulted in a significant number of deaths. A current example is the avian influenza that, according to World Health Organization (WHO) reports, has caused almost 61% fatality among the total number of cases reported from 2003 to September 2007 [1]. Another kind of threat is bioterrorism, where the threat is still biological but it is caused solely by humans [2]. In order to confront these situations, a number of disease surveillance systems have been developed that utilize a variety of indicators to detect possible disease outbreaks.

Groups of ‘vulnerable’ individuals such as patients suffering from chronic diseases (diabetes, chronic heart and renal failure), elderly people and infants could be considered as Sensitive Population Groups (SPGs). For example, diabetics are quite susceptible to infections that can be easily spread under certain circumstances and seem

¹ Corresponding author: Taxiarchis Botsis, Department of Computer Science, University of Tromsø, 9037 Tromsø, Norway. E-mail address: tbotsis@cs.uit.no

² This work was carried out during the tenure of an ERCIM “Alain Bensoussan” Fellowship Programme.

³ This work was partly funded by the Research Council of Norway, Project No: 174934.

to be affected more compared to normal subjects [3, 4]. Therefore, the development of electronic Disease Surveillance Systems (eDSSs) for SPGs is an issue that should be addressed. These systems could also aid health care professionals by providing them with early indications of any possible infectious disease outbreak.

Infectious diseases are a common cause of high blood glucose (BG) and hyperglycemic crisis in diabetics [5]. Generally, infection and any type of illness, surgery, or injury causes stress on the body that needs more energy to combat this situation. The body reacts by releasing counter-regulatory hormones that signal to the liver to release extra glucose in order to provide body with the requested energy. These hormones also inhibit the effect of insulin, and as a result there is over a certain period of time an increase in BG that is more difficult to control [6]. However, the detailed mechanism for BG elevation after infections has not been fully described yet [5].

Glycosylated hemoglobin (HBA_{1c}) is a reliable indicator of BG regulation, especially in diabetics. It represents the average BG level of the past 4 weeks approximately, strongly weighted toward the most recent 2 weeks. It is almost entirely insensitive to BG levels more than 4 weeks before. In non-diabetics, the formation, decomposition and destruction of HBA_{1c} reach a steady state with about 3.0% to 6.5% of the total hemoglobin, but most diabetics have a higher HBA_{1c} level. The actual HBA_{1c} level can be used as an indicator of the recent average BG level. Although HBA_{1c} varies among individuals with the same average BG, it is very stable for any given individual. Thus, a change of 1.0% in HBA_{1c} is definitely meaningful [7].

The target of this study was to demonstrate the need for developing DSSs for SPGs through the diabetics' case study. A simple system could be implemented using sensors for data collection, a simple network infrastructure for transmitting them to a central repository and an algorithm for data analysis; the same system could trigger alarms based on the analysis results. Given the fact that diabetics (especially type-1 diabetics) measure their BG daily, we claim that BG could be used as an indicator for the early detection of infections at the early stages of disease progression, i.e. during the incubation period. This period precedes symptoms onset which is the main time point that the subject is made aware of his/her infection. Specific indicators for other SPGs could be studied as well.

In order to establish a basis for further research it was considered necessary to find strong correlations between BG and infections. For this purpose the Diabetes Control and Complications Trial (DCCT) archives were used as the data source: DCCT was a full-scale multi-center clinical trial, which recruited 1441 type-1 diabetic patients and was conducted by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) [8]. Among the numerous detailed data that were collected, HBA_{1c} values and disease data were further assessed.

Methods

In DCCT patients were randomized into two arms: the conventional and the intensive therapy arm; data were collected mostly during the follow-up visit that was scheduled every three months.

In both arms adult males and females (age ≥ 18 years) were studied separately; the first 48 months of their involvement was the period of interest. A patient that was randomized to the conventional therapy arm was considered diseased if he/she reported a disease of more than five days at the follow-up visit. For the second arm, the

recording of even one disease incident was considered adequate. Moreover, for both arms, it was decided that the 6-month periods before and after disease documentation should be free of any other disease, and women pregnant at any point during or close to these periods was excluded from the analysis.

Accordingly, the HBA_{1c} values of the selected patients were assessed at specific time intervals: the follow-up visit of disease documentation (diseased value) and the 6-month periods before and after this visit (non-diseased values). The non-diseased HBA_{1c} values were averaged for both periods. According to DCCT protocol, HBA_{1c} values were collected every three months (at the follow-up visit) for the conventional therapy arm and every month for the intensive therapy arm. So, the non-diseased average was the result of two and five (the ones close to the diseased value were excluded) values correspondingly.

Sample distribution was tested for normality with the Shapiro-Wilk test, and the null hypothesis that the diseased HBA_{1c} values are equal to the non-diseased values was tested with paired t-test for both sexes and arms. All *p* values were based on two-sided testing, and differences were considered significant at *p*<0.05. All statistical analyses were performed with SPSS software (version 14.0 for Windows, SPSS Inc., Chicago, IL).

Results

The age range for all patients was 18-39 years; some other descriptive statistics are presented in Table 1. Figure 1 displays information on the HBA_{1c} grouped values (see abbreviations listed under the figure) distribution for both arms and sexes. The box plots and the outliers shown in this figure offer a quick overview for each group of values.

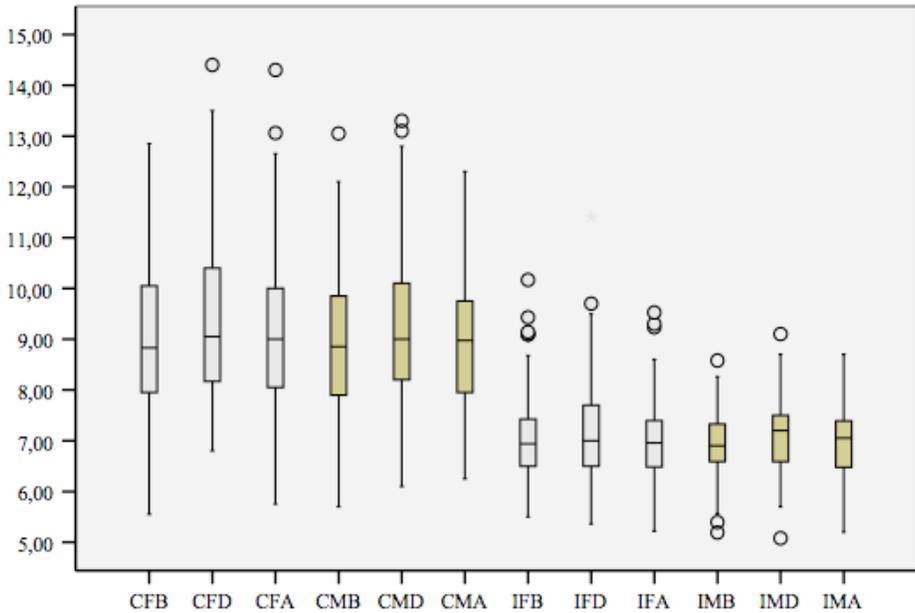
Table 1. Descriptive statistics for all patients

		Number	Age (years)	IDDM duration (months) *
Conventional Arm	Females	130	28.0 ± 5.9	79.0 ± 55.7
	Males	150	29.0 ± 5.6	64.8 ± 46.6
Intensive Arm	Females	101	29.0 ± 6.0	71.3 ± 50.9
	Males	95	28.8 ± 5.8	85.7 ± 51.6

*Insulin-dependent Diabetes Mellitus duration until patients' involvement

Table 2. Shapiro-Wilk test for Normality

HBA _{1c} Grouped Values	Conventional Arm		Intensive Arm	
	Statistic	<i>p</i> value	Statistic	<i>p</i> value
Female Diseased	0.970	0.006	0.941	<0.001
Female Average Before	0.987	0.234	0.943	<0.001
Female Average After	0.983	0.112	0.969	0.019
Male Diseased	0.986	0.135	0.992	0.819
Male Average Before	0.994	0.796	0.992	0.840
Male Average After	0.992	0.510	0.988	0.519



CFB: Conventional Female Before, CFD: Conventional Female Diseased, CFA: Conventional Female After, CMD: Conventional Male Diseased, CMB: Conventional Male Before, CMA: Conventional Male After, IFD: Intensive Female Diseased, IFB: Intensive Female Before, IFA: Intensive Female After, IMD: Intensive Male Diseased, IMB: Intensive Male Before, IMA: Intensive Male After.

Figure 1. Distribution of all HBA_{1c} grouped values for both arms and sexes.

The Shapiro-Wilk test showed that all sample distributions were normal except from females receiving intensive therapy (Table 2). In this case, all distributions of diseased and non-diseased average values appeared to be right-skewed. So, the

Table 3. The results of testing null hypothesis

		HBA _{1c}					
		Mean	SDev	95% CI	P		
Conventional Arm	Female	Diseased – Average Before	0.288	0.719	0.163	0.412	<0.001
		Diseased – Average After	0.273	0.851	0.126	0.421	<0.001
	Male	Diseased – Average Before	0.267	0.722	0.151	0.384	<0.001
		Diseased – Average After	0.225	0.681	0.115	0.335	<0.001
Intensive Arm	Female	Diseased – Average Before *	0.007	0.081	-0.009	0.023	0.342
		Diseased – Average After *	0.016	0.076	0.001	0.031	0.021
	Male	Diseased – Average Before	0.169	0.450	0.077	0.261	<0.001
		Diseased – Average After	0.115	0.563	0.001	0.230	0.049

* normalized values

logarithmic transformation (base-e logarithm) was used to correct the skew and normalize the distribution. It also appeared that the distribution for diseased females in the conventional arm was not normal either, but it was considered normal since there were only a few outliers and the dependent was quite normal, as appears in Figure 1.

The null hypothesis was tested with t-test and the results are presented in Table 3. All patients that were treated conventionally appeared to have increased HBA_{1c} levels at the follow-up visit they reported a disease incident compared to the average values for the 6-month periods before and after this visit. Particularly, females had a slightly greater increase in their diseased HBA_{1c} levels and a better recovery to lower non-diseased HBA_{1c} values compared to males.

The results for the intensive therapy arm were rather different. The null hypothesis was equally rejected for the male patients as well as for the patients of the conventional therapy arm. However, normalized diseased female values showed no significant difference compared to the normalized average values before but a significant decrease after the disease.

Discussion

Assuming that there could be some hormones pathways in adolescent subjects that could affect HBA_{1c} levels, only adults over 18 years were included in the analysis. The same assumption was made for pregnant women, who were excluded from this study. It could be argued that some other factors (physical activity, alcohol consumption, food intake) should be considered as well, but unfortunately there were no detailed recordings of them. Additionally, the recorded disease data was limited only to the number of disease incidents per patient for the intensive therapy arm, so in this case it was the only criterion for assessing whether a patient was ill or not. On the other hand, in the case of the conventional therapy arm, the days for each disease incident were also recorded. So, the criterion was modified: a reported disease incident that lasted less than 5 days was not considered important because it would be less capable of altering HBA_{1c} values.

The results showed that HBA_{1c} values rose after infection despite the tight BG control, and returned to non-diseased levels when the initial cause ceased. This was observed in all cases except the intensively treated females.

As mentioned above, infection causes release of counter-regulatory hormones and this results in BG elevation, but the detailed underlying mechanism has still to be determined. Moreover, the exact timing of BG elevation during the disease progression requires more clarification. It could be expected that both mechanism and timing follow quite different patterns in SPGs case and, especially, in diabetics who were the target group of this study. Particularly, a number of relevant questions could arise such as:

- Is it possible to observe a meaningful increase in HBA_{1c} after infections?
- How well is BG regulated with insulin intake during disease episodes?
- Is BG a reliable indicator since it is affected by many other factors?
- What are the different patterns of BG elevation per individual and infection?

Our study indicated a meaningful increase in HBA_{1c} after infections but this should serve only as a basis for further research. Ukaparol et al. [9] have partially answered the same question by their study in animals. Regarding the second question, there are findings that BG is increased in patients with postoperative infections even with tighter

glucose control [10, 11] (our analysis resulted in similar findings), but more targeted research is needed. Despite the fact that BG is affected by many other factors, such as food intake, insulin, physical activity and alcohol, the DCCT data analysis indicated that BG cannot be easily regulated after infection even when the patient is closely monitored for the other factors. A potential direction for future investigation is stated through the last question which is also our idea for the next steps in the field.

Electronic disease surveillance systems can offer early warnings for a disease threat and/or possible outbreaks. An eDSS for diabetics could employ biological sensors for collecting BG data and transmit them to a central data repository and their Electronic Health Record. Then an algorithm could analyze the incoming values in respect to each subject's profile, identify epidemic threats for a bigger group of diabetics (or the general population) and trigger alarms if necessary. Naturally, the use of BG data and the implementation of similar systems for non-diabetics should be studied as well.

As stated above, the results of our study indicate that before developing an eDSS for diabetics more research is required. So, the next step is a pilot study with diabetics. The target of this study will be the collection of the necessary data that will be used in studying the different patterns and physiology mechanisms after infections in diabetics. A future plan is to investigate whether it is possible to apply the same strategies for the general population. Apparently the diabetics' case study could act as the force for electronic disease surveillance not only for the general population but also for other sensitive population groups. However, a dedicated study should be performed per case in order to derive safe conclusions.

References

- [1] Cumulative Number of Confirmed Human Cases of Avian Influenza A/(H5N1) Reported to WHO. See http://www.who.int/csr/disease/avian_influenza/country/cases_table_2007_09_10/en/index.html (Last checked 01 October 2007).
- [2] D.A. Henderson, T.V. Inglesby, T. O' Toole, eds. *Bioterrorism: Guidelines for medical and public health management*. AMA press, 2002.
- [3] P.C. Reading, J.A. Erika, C. Crouch, E.M. Anders Increased susceptibility of diabetic mice to influenza virus infection: Compromise of collectin-mediated host defense of the lung by glucose? *Journal of Virology* **72**; **8** (1998), 6884–6887.
- [4] E.H. Baker, D.M. Wood, A.L. Brennan, N. Clark, D.L. Baines, B.J. Philips. Hyperglycemia and pulmonary infection. *Proceedings of the Nutrition Society* **65** (2006), 227-235.
- [5] B. Bharat. Diabetes and infection. *Academy Journal of Applied-Basic Medical Sciences* **5** (2003).
- [6] J. Kohn. Infection can raise blood glucose. *Abbott Diabetes care – Ask the expert*. See http://www.abbottdiabetescare.com/adc_dotcom/url/questionAnswerProfile/en_US/40.20:20/question_answer/question_answer/QuestionAnswer_00115.htm (Last checked 01 October 2007).
- [7] What's HBA_{1c} and what's it mean? *Diabetes FAQ: BG monitoring*. See <http://www.faqs.org/faqs/diabetes/faq/part2/section-9.html> (Last checked 01 October 2007).
- [8] *Diabetes Control and Complications Trial (DCCT)*. See <http://diabetes.niddk.nih.gov/dm/pubs/control/> (Last checked 01 October 2007).
- [9] N. Ukarapol, R.E. Begue, J. Hempe, H. Correa, R. Gomez, A. Vargas. Association between helicobacter felis–induced gastritis and elevated glycated hemoglobin levels in a mouse model of type 1 diabetes. *The Journal of Infectious Diseases* **185** (2002), 1463–1467.
- [10] B.O. Kim, S.W. Kim, S.M. Lee, H. Shin. Relationship of glucose control and wound infection in diabetics after lumbar spine surgery. *Journal of Korean Neurosurgery Society* **37** (2005), 44-47.
- [11] M. Sander, C. Heymann, V. Dossow, C. Spaethe, Q.F. Konertz, U. Jain, C.D. Spies. Increased interleukin-6 after cardiac surgery predicts infection. *Anesthesia & Analgesia* **102** (2006), 1623-1629.

Clinical Processes in an Innovative Vascular Surgeon Community. Implications for Workflow Modeling

Berit BRATTHEIM ^{a,1}, Andreas R. SEIM ^b and Arild FAXVAAG ^c

^a *The Norwegian EHR research centre (NSEP),*

Institute of neuroscience, Faculty of medicine, NTNU and

The Faculty of Food Science and Medical Technology,

Sør-Trøndelag University College

^b *Department of Productivity and Project Management, SINTEF*

^c *NSEP, Faculty of medicine, NTNU Trondheim, Norway*

Abstract: *Objective:* To identify factors influencing variations in clinical work in the care of patients with abdominal aortic aneurism. *Method:* Ethnographic observations of 26 meetings between surgeons and patients in two community hospitals and one university hospital. Observations data were abstracted into scenarios that describe the typical clinical workflow. Characterizations of features of the scenarios were performed. *Results:* When comparing the university hospital and the community hospitals we find large variations in patient trajectories, and in the relation between actors' and roles. *Conclusion:* Given a clinical domain distinguished by an unrelenting search for new and improved surgical techniques, workflow system requirements should reflect that healthcare planning not only is conducted with the purpose of providing care but also with purpose of developing new or maintaining existing surgical skills.

Keywords: workflow systems, clinical process, cardiovascular, decision-support

Introduction

Clinical practice guidelines are systematically developed statements that describe how healthcare should be enacted to comply with the principles of Evidence-based medicine (EBM). It is widely assumed that implementation of clinical practice guidelines will improve both quality and efficiency of care. Accordingly, numerous initiatives have been taken to develop and implement healthcare information systems aiming to facilitate the delivery of guideline-based care [1-4]. As part of developing requirements for such information systems, studies have applied methods for observing complex care situations to develop frameworks and concepts for characterizing the specific situations [5-7]. Such observations have, however, revealed that there are a number of problems

¹ Corresponding author: Berit BRATTHEIM, The Norwegian EHR Research Centre, Medical technical research centre, N-7489 Trondheim, Norway, E-mail: Berit.J.Brattheim@ntnu.no

with guidelines. For instance, that clinicians' practice often deviates from the actions suggested in guidelines, and these deviations are often well justified [5,8-9]. Accordingly, many healthcare workflow system implementations have failed when confronted with the complex reality at a clinical ward [4].

In this paper we present preliminary data from a study of clinical practice related to monitoring of patients with severe abdominal aortic aneurism (AAA) to identify patients eligible for surgical intervention. As of 2007, the conventional method of open surgery for AAA is about to be replaced with a more advanced endovascular method (insertion of an aortic prosthesis (EVAR)). Evaluation of AAA for inclusion to stent-graft surgery as well as the actual insertion of the prosthesis requires the coordinated action of a multidisciplinary care team consisting of both radiologists and vascular surgeons [10-13]. The EVAR method is now accepted as superior to conventional open surgery [13]. Being a novel, technically demanding and technologically more advanced method, it is mainly practiced by vascular surgeons at university hospitals. In this study, we compared work practice at one university hospital with that at two community hospitals. We present data that illustrates variations in clinical practice, analyze which factors might influence or induce variation, and briefly discuss the implications for healthcare workflow systems.

1. Materials and methods

Healthcare scenario: Meetings between patients and vascular surgeons at one university hospital and two community hospitals owned by Central Norway Regional Health Authority.

Study design: Ethnographic observations of encounters between patients and surgeons. The contexts for data collection were two clinical situations: a) the monitoring of AAA patients potentially eligible for vascular surgery at scheduled visits, b) the decision making process leading to a choice on which surgical technique to use.

Data collection: A semi-structured observation protocol was used for recording information, including patient demographics, actors, roles, information sources, events and decisions. Narratives describing each observation were written based on the observable data and statements from the participants. These were abstracted into scenarios describing the typical course of events for the chosen situations. Each scenario was considered as an input-output process as suggested in reference [6], and a conceptual framework was used as a starting point for identifying typical scenario features [6,7].

Ethical considerations: The study was approved by the regional committee for medical research ethics and the Norwegian social science data services. Access to the field was obtained from the physicians in charge at both the radiological ward and outpatient vascular surgery ward at all three hospitals. Patients eligible for inclusion were informed about the study by letter from a coordinator or the head physician. Only patients who signed the informed consent form were recruited. Signed informed consent was also obtained from the physicians performing the scheduled visits.

2. Results:

2.1. Patient demographics

26 meetings between patient and vascular surgeon were observed. These included 7 female and 19 male patients: age ranging from 57 to 94 years old. Characterizing features for the *majority* were: a) the AAA was discovered by coincidence as a secondary finding during examination of another disease, b) AAA diameter ranged from 5.0 to 6.5 cm, c) a suffering of heart- and/or lung diseases.

2.2. Characteristics of patient trajectories

As stated in the materials and methods section we recorded data like actors, roles, tasks and decisions. Based on the narratives we outlined descriptions of events in a) the principal AAA patient monitoring trajectory, and b) the decision-making trajectory. Examples of both types of events are given in Table 1. Both were characterized by the surgeon performing a risk-assessment, the use of image report as a decision basis, information to the patient related to their particular aneurysm as well as how the forthcoming trajectory would transpire.

Table 1. Example of patient trajectories at the university hospital and a community hospital

Event A1: Watchful waiting: university hospital.

81-year-old-man referred for monitoring of AAA. The ultrasound report showed an aneurysm diameter of 4.7 cm. At the clinical visit, one surgeon from the vascular surgeon team assessed the patient's clinical status, performing a risk-assessment. The patient was informed about the size/growth of the aneurysm since last control as well as being advised about self-care. The surgeon considered the aneurysm to be stable with a diameter well below the treatment threshold. He informed the patient that risk of treatment was greater than the risk of no-treatment. A new consultation was scheduled within 6 months – this time with a CT exam to get an overview of the anatomic characteristics.

Event A2: Watchful waiting: community hospital.

65-year-old-man referred for monitoring of AAA. The ultrasound report showed an aneurysm diameter of 4.6 cm, a growth of 0.5 cm since the last visit which had taken place one year earlier. At the clinical visit, the vascular surgeon assessed the patient's clinical status, performing a risk-assessment. The patient was informed about the size/growth of the aneurysm since last control as well as being advised about self-care. The patient wanted to be considered for EVAR treatment. The surgeon explained that the threshold limit for treatment was a diameter of 5.2-5.5 cm and that the patient was a bit too young for the EVAR option. He informed the patient that risk of treatment was greater than the risk of no-treatment. In agreement with the patient, a new consultation was scheduled combined with an ultrasound examination.

Event B1: Patient trajectory at university hospital.

67-year-old man who was referred for assessment on treatment. Before the clinical visit, a CT exam was undertaken for an evaluation of suitability for EVAR. Decision on EVAR was done in a meeting between intervention radiologists and vascular surgeons, and EVAR was recommended. At the clinical visit, one surgeon at the vascular surgeon team assessed the patient's clinical status, performing a risk-assessment. The surgeon informed about the EVAR option and discussed the option with the patient. As both actors agreed on the EVAR alternative, the next step was to schedule the patient for EVAR within 1-3 month.

Event B2: Patient trajectory at community hospital.

73-year-old man who was referred for assessment on treatment. A CT exam was performed to get an overview of the anatomic characteristics. At the clinical visit the vascular surgeon assessed the patient's clinical status performing a risk-assessment. The surgeon informed about the treatment alternatives. It was decided to proceed with an EVAR assessment. The surgeon would then make a request to the vascular surgeon in charge at the university hospital who would perform a relevance assessment and request the radiologist to assess the CT-exam from the community hospital as a decision basis for EVAR suitability.

2.3. More complex decision making trajectories at community hospitals

The principal AAA patient-monitoring trajectory was quite similar at all three hospitals. However, the decision making trajectories were different (figures 1 and 2). At the community hospitals, whenever a patient eligible for EVAR was identified, another partial information routing path was triggered. The local actor had to write a referral letter to the vascular surgeon in charge at the university hospital. If urgent, the local actor would in fact also consult the person concerned by phone. Then, the university hospital surgeon performed a relevance-assessment and requested the radiological department for judgement of suitability. In the case of EVAR, the patient responsibility would be transferred from the local actor to the vascular surgeon team at the university hospital. Thus, our findings indicated a *more complex decision making trajectory at the community level* with respect to the local actor enacting multiple roles, other actors to be involved, increased information exchange, an extra decision node as well as a change in patient stream.

2.4. Variations in the use of imaging technologies and in the practice of rules for inclusion of patients to endovascular surgery:

Looking at the practices of ordering imaging analyses, we found differences between the university hospital and the community hospitals. Surgeons at the former tended to order more CT exams as part of the screening process. Surgeons at community hospitals however preferred to measure the diameter of the aneurysms by use of ultrasound, stating the necessity of making all recordings with use of the same imaging technique. Another example of variation was the practice of the rule of a lower age limit for patients eligible for EVAR. The university hospital agreed on an informal rule of a lower age limit of 60, while the local vascular surgeons seemed not to recommend EVAR until the age of 65-70.

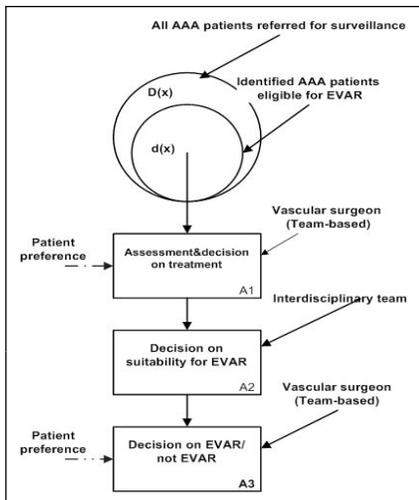


Figure 1. Patient trajectory at university hospital

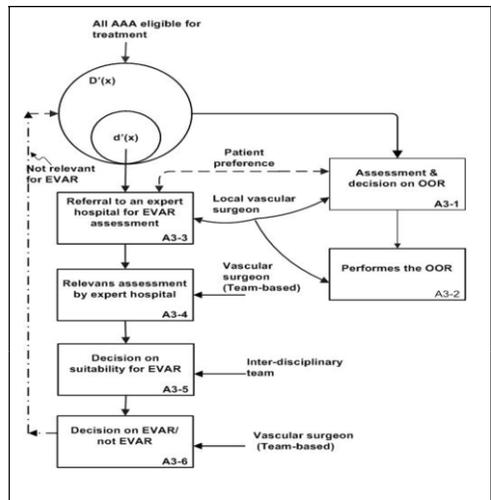


Figure 2. Patient trajectory at community hospital

2.5. Same healthcare actor - different roles

To obtain characterizing features, we used the scenarios to extract information on the involved actors and their roles. For the surgeon, two main roles were identified: the role of being a vascular surgeon and the role of scheduling and planning at the department level. As for the latter, this can be exemplified with the maintenance of patient stream related to allocating patients to the hospital and reporting to registries. Regarding the role of vascular surgeon, this encompassed a role as care provider as well as that of a person responsible for developing new or maintaining existing surgical skills. The university hospital had a team-based workflow with several surgeons. The roles were maintained either collectively or individually. In contrast, the community level was characterized by an individual aspect. The local actor kept both the role as hospital planner and that of vascular surgeon. However, the introduction of the EVAR method influenced both roles: now the local actor was assigned the task of identifying patients eligible for EVAR thus leaving a reduced number of AAA patients to be treated by the conventional surgery method. To summarize, the university level had *many actors – many roles* as opposed to the community level which had *one actor-many roles*.

3. Discussion:

Our findings demonstrate a more complex decision making trajectory at the community hospitals, variations in actor's role and practice behaviour. The main findings are related to the actor–role aspect. At the community level, there is only one actor enacting multiple roles: The actor is both an expert in vascular surgery and a gatekeeper responsible for recruiting patients to the community hospital. The EVAR method continuously improves, and the proportion of AAA patients eligible for EVAR is increasing accordingly. The more patients being routed to EVAR, the fewer patients are left for the conventional open surgery by the local surgeon. Personnel skills are also important for the task of identifying candidates eligible for EVAR. It is important to avoid that the EVAR alternative at one hospital can be offered to patients not being selected as eligible candidate at another hospital. In contrast, the university hospital is team-based with several actors to maintain multiple roles as a collective function.

The increased complexity in patient trajectory at community level led to a less streamlined patient flow. Throughout the trajectory, the responsibilities of the personnel changed, and actors outside the community hospital had to be engaged. Taken together, these developments challenged the organizational principle of continuity of care.

The observed variations in clinical practice because of the introduction of more advanced treatment alternatives illustrate the complexity of the clinical process. What is being produced is not only healthcare but also health personnel that master a set of skills. The illustrated variations in work practice may have implications for the architecture of healthcare workflow systems and other systems tailored to the healthcare domain. There is obviously a need to reduce inappropriate variations in healthcare practice. Because of the rapid development and introduction of improved treatment methods, such systems must be highly flexible. To be able to assign healthcare tasks to the proper actors, workflow systems must distinguish between actors that possess a particular skill and actors that not possess but are in the position to

develop the particular skill. To conclude, we believe that both the production of care and the production of skills deserve the support from a workflow system.

This study has several limitations. The sample size was small and we cannot exclude the presence of research bias with regards to the observations. We could have sought to verify our findings by collecting retrospective data from the physician's medical record notes of the described events as well as from the registered procedure- and diagnosis codes in the National Patient Register. However, this method would require special approval from the National Health Authority. We currently are extending the study by using a mixed-method approach thus supplementing the empirical data with interviews of health key personnel and patients. Data from these studies will be presented in separate studies.

4. Conclusion

Being a domain in rapid development, we found that the introduction of new treatment method introduced more complex decision making trajectory at the community hospitals than at the university hospital. Our findings also indicated variations in actor's role and practice behaviour. This has implications for the modelling of workflow and the workflow systems.

References

- [1] A.X Garg, N.K. Adhikari, H. McDonald, M.P. Rosas-Arenello, P.J. Devereaux, J. Beyene, J. Sam, R.B. Haynes, Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes, *JAMA* **293**(10) (2005), 1223-38.
- [2] S. Quaglini, M. Stefanelli, A. Cavallini, G. Micieli, C. Fassino, C. Mossa, Guideline-based careflow systems, *Artificial Intelligence in Medicine* **20**(1) (2000), 5-22.
- [3] S. Panzarasa, S. Maddè, S. Quaglini, C. Pistarini, M. Stefanelli, Evidence-based careflow management systems: the case of post-stroke rehabilitation, *Journ. of Biomedical Informatics*, **35**(2) (2002), 123-139.
- [4] B. Chaudhry, J. Wang, S. Wu, M. Maglione, W. Mojica, E. Roth, S.C. Morton, P.G. Shekelle, Systematic Review: Impact of Health Information Technology on Quality, Efficiency, and Costs of Medical Care, *Annals of Internal Medicine*, **144**(10) (2006), 742-752.
- [5] A. Kumar, B. Smith, D.M. Pisanelli, A. Gangemi, M. Stefanelli, Clinical guidelines as plans: an ontological theory, *Methods of Information in Medicine*, **45**(2) (2006), 204-10.
- [6] I.D. Sørby, L. Melby, Ø. Nytrø, Characterizing cooperation in the ward: framework for producing requirements to mobile electronic healthcare records, *Int. Journal of Healthcare Technology and Management* **7** (6), (2006), 506-521.
- [7] I.D. Sørby, Ø. Nytrø, T. B. Røst, Empirical Grounding of Guideline Implementation in Cooperative Clinical Care Situations. Proc. from workshop: AI techniques in healthcare: computerized guidelines and protocols, Riva del Garda, Italy, 29.08.06.
- [8] T.J. Jordan, Understanding Medical information: A User's Guide to Informatics and Decision making. *McGraw-Hill medical Publishing Division*, 2002.
- [9] P. Taylor, From Patient Data to Medical Knowledge, *BMJ*, *Blackwell Publishing*, 2006.
- [10] M. Al-Omran, S. Verma, T.F. Lindsay, R.D. Weisel, Y. Sternbach, Clinical Decision Making for Endovascular repair of Abdominal Aortic Aneurysm, *Circulation – American Heart Association*, **110** (2004), 517-523.
- [11] R.M. Greenhalgh, L.C. Brown, G.P. Kwong, J.T. Powell, S.G. Thomson, Comparison of endovascular aneurysm repair with open repair in patients with abdominal aortic aneurysm (EVAR 1): Randomized controlled trial, *Lancet* **364**(9437) (2004), 843-8.
- [12] EVAR trial participants: Endovascular aneurysm repair and outcome in patients unfit for open repair of abdominal aortic aneurysm (EVAR 2): Randomized controlled trial, *Lancet* **365**(9478) (2005), 2187-92.
- [13] "Paradigmeskift ifrån open till endovaskulär aneurysmkirurg?"
<http://www.svenskkirurgi.se/skf/svkir/06-4/EVAR.htm>, 2006-09-12 (in Swedish)

Medical Knowledge Representation System

David BUCHTELA, Jan PELESKA, Miroslav ZVOLSKY, Jana ZVAROVA
EuroMISE Center and Center for Biomedical Informatics, Department of Medical Informatics, Institute of Computer Science AS CR v.v.i., Pod Vodarenskou vezi 2, 182 07 Prague 8, Czech Republic

Abstract. The aim of this article is to present a design of a Medical Knowledge Representation System (MEKRES). The system automatically offers relevant formalized knowledge by extended GLIF (Guidelines Interchange Format) models to participants (patient, , physician, operator, ...) on the basis of acquired data. This selection algorithm is based on key attributes and cooperation with knowledge ontologies.

Keywords. Knowledge Representation, GLIF model, Ontology

Introduction

Nowadays the conceptual knowledge (concepts, targets, actions,...) plays more prominent role. Conceptual knowledge is related to given task but expressed on higher level of abstraction and relatively independent of the form which is used by a knowledge system or expert. These factors are related to the creation and development of structured methodologies which support model based knowledge mining and using.

The aim of this article is to present a design of a Medical Knowledge Representation System which is based on GLIF models of branch knowledge. The GLIF models are extended by key attributes. The MEKRES offers the participants (patient, physician, operator,...) relevant model(s) based on acquired data.

1. GLIF model

The GLIF (*Guideline Interchange Format*) model is a result of collaboration among Columbia University, Harvard University, McGill University and Stanford University. The main goal of GLIF was to enable sharing of guidelines among institutions and across computer applications [3].

GLIF specifies a process-oriented model for guidelines representation and syntax for guidelines utilization in software systems as well as for their transport. GLIF guidelines are mostly given as a flowchart representing a temporarily ordered sequence of steps. The nodes of the graph are guideline steps and edges represent continuation

from one step to the other one (see fig.1). Guideline steps are an *action step*, *decision step*, *branch* and *synchronization steps* and a *patient state step*.

- **Action steps** specify clinical actions that are to be performed. It can be an application of some therapy, carrying out some examination or measurement etc. Action step also may name sub-guidelines (subgraph), which provide a detail for the action.
- **Decision steps** are used for conditional branching. There are two kinds of decision steps: **Case step** is used, when branching is determined by evaluation of defined logical criteria based on data items. **Choice step** is used when the decision cannot be precisely specified in guidelines themselves and decision should be made by the user.
- **Branch** and **synchronization steps** enable concurrence in the model. Guideline steps that follow branch step can be performed concurrently. Branches with root in branch step eventually converge in a synchronization step. In this step all branches are synchronized after evaluation of synchronizing condition.
- **State step** characterizes a surveyed object state after execution of the previous step or in the beginning of the model.

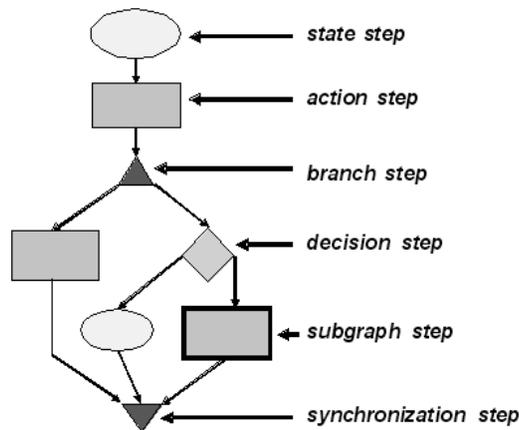


Figure 1. Graphical symbols of GLIF model

1.1. Criteria of conditions

The decision step specifies several criteria of condition for each decision option:

- The **strict-in** criterion is used to specify a decision condition that could be computed automatically (for example if systolic blood pressure is 130 or greater). If a strict-in is true then the control flows to the guideline step that is specified by that decision option's destination.
- The **strict-out** criterion is analogous to an absolute contraindication (for example if a patient is gouty he could not be cured by thiazides diuretics). If a strict-out is true then the decision option's destination is forbidden.
- The **rule-in** criteria rank a choice as the best among several options. For example, when there are competing diagnoses for a disease, a pathognomonic

condition would be a rule-in for the disease. This criterion is analogous to conditions favouring the use (indications).

- A **rule-out** takes precedence over rule-in when ranking options. If an option contains both a rule-in criterion and a rule-out criterion, and both are evaluated as true, then that option should be the last choice. This criterion is analogous to contraindications.

The strict-out criterion is evaluated at first. If strict-out criterion is evaluated as true the rest of the criteria is not evaluated. This option is forbidden. In the opposite case the strict-in criterion is evaluated. If the strict-in criterion is false too, the rule-in and rule-out criteria are evaluated. The ranking of rule-ins and rule-outs is left to the users who may use their clinical judgment or may develop their own ranking schemes.

1.2. GLIF implementation

GLIF model is graphical so it is necessary to code it in a formal language. These formats are available for implementation of GLIF model and its parameters:

- **GELLO** (*Guideline Expression Language*) – the object oriented query and expression language for decision support [3].
- **RDF** (*Resource description framework*) – the built-in form of saved knowledge base in a Protégé system [2]. The Protégé is a development environment used for knowledge system design. The environment enables knowledge ontology construction, data inserting and user formatting of input and output forms.
- **XML** (*eXtensible Markup Language*) – The encoded model consists of a sequence of guideline steps. Some attributes of a guideline step contain next guideline steps. It enables sequential representation of a graph structure in the guideline language.

2. Knowledge ontology

Knowledge ontologies follow the development in an artificial intelligence in an area of knowledge representation. Ontologies are thoroughly taken in logic theories and their link to real objects (instances) is relatively opened. The ontology is primarily used to describe concepts (classes) and not to describe facts about concrete objects. Classes (concepts) and relations are systematically defined by a formal language [4].

The *class* is the base of the knowledge ontology and describes a set of objects. In some formalisms the *class* corresponds to a *concept* or a *category* and closely corresponds to a *frame* which is the base construction of any artificial intelligence systems. The ontology classes don't contain procedural methods in opposite to classes in object-oriented models.

On a set of classes there is defined a hierarchy (taxonomy). The philosophical view of ontology sometimes requires a strict tree structure but all of the main ontology languages support multiple heredity.

The *individuum* corresponds to a concrete object of the real world and is the opposite of the class in a way. The individuum can be inserted into the ontology without a link to any class.

3. Medical Knowledge Representation System

3.1. Knowledge Representation Model

Knowledge Representation Model (KREM) of the whole system is based on a branch knowledge formalisation through the use of the GLIF model. The formalisation process, i.e. construction of the graphic GLIF model of knowledge contained in free text and the model coding into the formal language (XML)[1], is illustrated in figure 2.

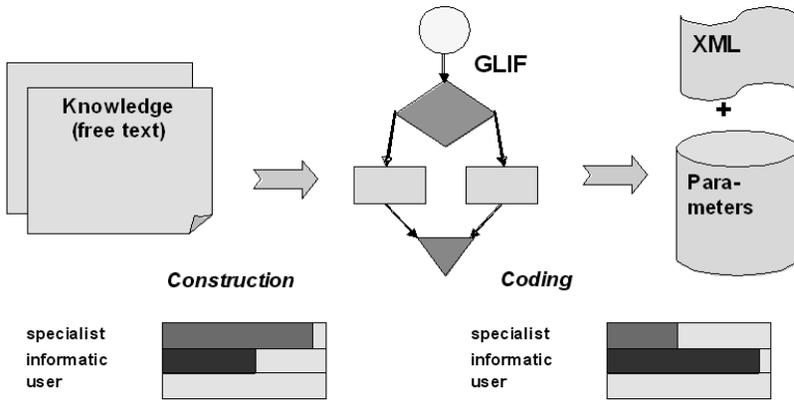


Figure 2. Process of GLIF model construction and implementation

In the stage of a GLIF model construction from a free text it is important to find a logical and process structure of knowledge, all fundamental parameters and their interrelationships. The result of this stage is a graphic GLIF model corresponding to the knowledge in the text. The construction stage is the most important and difficult of all stages.

In the stage of GLIF model coding the graphic model of knowledge is coded into XML. Some steps of the encoded model contain description of next steps (next options). It enables sequential representation of a graph structure [1].

Table 1. Key attributes of knowledge model

Key attribute	Attribute description
branch	Branch described by the GLIF model – e.g. cardiology
BID	Branch identification – e.g. International Classification of Diseases (ICD)
user	User of system to whom the GLIF model is primary determined (patient, physician, operator, ...)
status	The GLIF model validity
key	List of keys described by the GLIF model – e.g. blood pressure, diabetes, ...
key_name	Name of the key
key_weight	Weight of the key – the GLIF model description rate of the key

The resulting GLIF model is extended by key attributes. The list of the key attributes is in the table 1. These key attributes are used in selection algorithm and they are coded in XML along with the GLIF model.

3.2. Selection algorithm

The principle of the Medical Knowledge Representation system (MEKRES) and the algorithm of the relevant GLIF model selection is illustrated in figure 3.

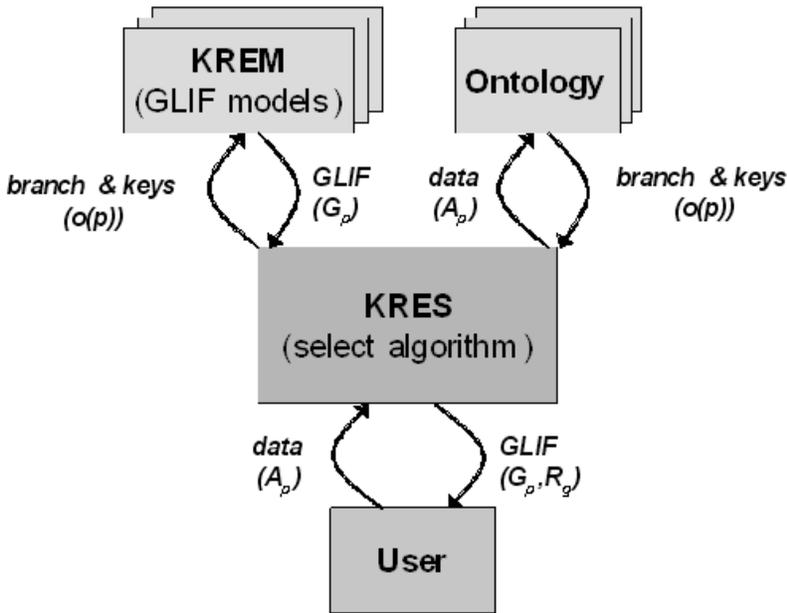


Figure 3. Medical Knowledge Representation System

The selection algorithm can be described subsequently:

- For every specific participant (user, patient, physician, operator, ...) p and his attributes A_p a set of all branches (areas) $\exp(B)$ which corresponds to participant state (attributes A_p) is determined. This function o is determined using the knowledge ontology:

$$o: A_p \longrightarrow \exp(B), \text{ where } B \text{ is a finite set of all branches.}$$
- For each branch $B \in o(p)$ a subset $K(p, B)$ of recognized attributes (keys) of the branch B and affected by attributes of the participant p is determined.
- Models G_p are chosen from a finite set of all GLIF models M thus:

$$G_p = \{g \in M : g \in K(p, B) \cap \text{key_weight}(\text{branch}(g), P(g)) \neq 0\},$$
 where each GLIF model contains attribute *branch* and a finite set P of keys (attribute *key_name*) and their weights (attribute *key_weight*).
- For each model G_p a general aggregate operator r is defined:

$r: \{G_p\} \rightarrow \mathfrak{R}$, where \mathfrak{R} is the set of real numbers.

The operator r can be for example defined as $R_g = \sum_{k \in P(g)} key_weight(k)$. The participant is then offered the GLIF model with the highest relevance value $R = \max(R_g)$ or a list of models ordered by the value of R_g .

4. Conclusion

Designed medical knowledge representation system is based on the GLIF model which is the universal method of modelling of mainly procedural-oriented knowledge. The system offers a relevant GLIF model or an ordered list of models from a list of available formal models on the basis of participant attributes. Additional information describing the concrete situation gives the participant possibility of better decision.

Acknowledgement

The work was partially supported by the project 1M06014 of the Ministry of Education of the Czech Republic and by the research institutional plan AV0Z10300504 of the Institute of Computer Science AS CR v.v.i.

References

- [1] Buchtela D., Peleska J., Vesely A., Zvarova J. Method of GLIF model Construction and Implementation, *Stud Health Technol Inform* **116** (2005), 779-784.
- [2] Gennari J.H., Musen M.A., Fergeson R.W., Grosso W.E., Crubezy M., Eriksson H., Noy N.F., Tu S.W.: The evolution of Protégé: an environment for knowledge-based systems development, *Int J Human-Comput Stud* **58**, (2003), 84-123.
- [3] Ohno-Machado L., Gennari J.H. et al., The GuideLine interchange format: a model for representing guidelines, *J Am Med Inform Assoc* **5** (1998), 357-372.
- [4] Uschold M., Gruninger M., *Ontologies: principles, methods and applications*. *Knowl Eng Rev* **11** (1996), 93-136.

An electronic registry for physiotherapists in Belgium

Ronald BUYL^{a*}, Marc NYSSSEN^a

^a*Department of Biostatistics and Medical Informatics, Faculty of Medicine, Vrije Universiteit Brussel, Belgium*

Abstract. This paper describes the results of the KINELECTRICS project. Since more and more clinical documents are stored and transmitted in an electronic way, the aim of this project was to design an electronic version of the registry that contains all acts of physiotherapists. The solution we present here, not only meets all legal constraints, but also enables to verify the traceability and inalterability of the generated documents, by means of SHA-256 codes. The proposed structure, using XML technology can also form a basis for the development of tools that can be used by the controlling authorities. By means of a certification procedure for software systems, we succeeded in developing a user friendly system that enables end-users that use a quality labeled software package, to automatically produce all the legally necessary documents concerning the registry. Moreover, we hope that this development will be an incentive for non-users to start working in an electronic way.

Keywords. Electronic Health record (EHR), Physiotherapy, Registry, XML

Introduction

During the past decade the healthcare industry has evolved from paper-based storage of clinical data into the digital era. Electronic healthcare records play a crucial role to meet the growing need for integrated data-storage and data communication [1,2]. Since 1998 for the software systems for general practitioners and since 2002 for software systems for physiotherapists (and other healthcare providers) the Belgian Ministry of Health, Food Chain Safety and Environment is issuing quality labels to systems that meet the well specified quality criteria, with an ultimate goal to improve the electronic record keeping for all healthcare workers in Belgium. In this context of improved record keeping, the KINELECTRICS project must be situated.

According to a new law issued on December 7th, 2005, physiotherapists (but also nurses and speech therapists) are required to keep an electronic version of the registry, which contains all physiotherapeutic acts, starting from January 1, 2007. Up until that day, a paper version of the registry had to be created every month. This hand written (or printed out) paper version had to be provided in a bind and page numbered booklet containing all the acts carried out by the physiotherapist, sorted by date. This

* Corresponding Author: Ronald Buyl, Departement of Biostatistics and Medical Informatics, Faculty of Medicine, Vrije Universiteit Brussel, Laarbeeklaan 103, B-1090 Jette (Brussels); E-mail: rbuy1@vub.ac.be

physiotherapist's registry can not be confused with the classical form of registries, like e.g. national cancer registries [9], where the goal is to collect, consolidate and classify data from different hospitals or institutes and link them to data within other registries or medical record systems. The physiotherapist's registry is a legal obligation that can be compared with the summary records accountants have to provide, which contain summarized information about the produced invoices over the past year.

The goal of the KINELECTRICS project was to come up with a suitable electronic version that could replace the old paper one. But electronic data storage is threatened by the same basic hazards as paper storage. Data can disappear; integrity can be lost, together with the ability to understand its content.

The project succeeded, after an in dept analysis of the legal context and a with particular attention to the reporting habits within the field of physiotherapy, to come up with a new structure for the electronic registry. This new structure was tested during the certification procedure of November 2006, by means of six additional test criteria and a specially adapted test scenario.

1. Material and methods

The KINELECTRICS project consisted of three different phases. In a first phase, taking into account all legal aspects, a model for the normal flow of registered data as well for physiotherapist using a quality labeled software package as for those without, was worked out. In this stage the data-structure was firmly defined. In the second phase, a study was performed on two key aspect of the project namely the traceability and the inalterability of the registered data. In a third and final phase the test criteria and scenarios were developed, including the elaboration of a basic control mechanism that was used during the certification procedure and can be used as a basis for more advanced inspection tools.

1.1. Phase 1: flow of registered data and data-structure

According to the current legal requirements, the content of the electronic version of the physiotherapy registry does not differ too much from the paper-based format. Following items were considered to be sufficient to be able to verify the process of the acts of a physiotherapist (for every act performed):

- Date of the act
- Name of the patient
- Nomenclature number of the act

As a data-structure for the registry we opted for an XML-message. XML has already proven its benefits as an electronic language for the communication of medical data [3,4,5]. Furthermore Kmehr-Bis¹, an XML messaging schema, is already used as a standard in Belgium for the exchange of medical documents, such as discharge letters, lab results and medical prescriptions. An example of a structured Kmehr message for the registry is provided in figure 1.

¹ Kmehr-BIS: Kind Messages for Electronic Health Records, Belgian Implementation Standard: <http://www.chu-charleroi.be/kmehr/htm/kmehr.htm> [6]

```

<kmehrmessage xmlns="http://www.health.fgov.be/telematics/kmehr/schema"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.health.fgov.be/telematics/kmehr/schema ..\xsd\kmehr.xsd">
  <header>
    <standard>
      <cd S="CD-STANDARD" SV="1.0">20021205</cd>
    </standard>
    <id S="ID-KMEHR" SV="1.0">***.***</id>
    <date>2006-07-31</date>
    <sender>
      <hparty>
        <id S="ID-HCPARTY" SV="1.0">1234567890</id>
        <cd S="CD-HCPARTY" SV="1.0">persphysiotherapist</cd>
        <firstname>Dirk</firstname>
        <familyname>Masseur</familyname>
      </hparty>
    </sender>
    <recipient>
      <hparty>
        <id S="ID-HCPARTY" SV="1.0">1234567890</id>
        <cd S="CD-HCPARTY" SV="1.0">persphysiotherapist</cd>
        <firstname>Dirk</firstname>
        <familyname>Masseur</familyname>
      </hparty>
    </recipient>
    <sha>
      <id S="ID-SHA-256" SV="1.0">1</id>
    <checksum>e3b0c44298fc1c149afb4c8996fb92427ae41e4649b934ca495991b7852b855</checksum>
    </sha>
  </header>
  <folder>
    <id S="ID-KMEHR" SV="1.0">1</id>
    <patient>
      <id S="ID-PATIENT" SV="1.0" />
      <firstname>A</firstname>
      <familyname>EEN</familyname>
      <sex />
    </patient>
    <transaction>
      <id S="ID-KMEHR" SV="1.0">1</id>
      <cd S="CD-TRANSACTION" SV="1.0">idefix_recording</cd>
      <date>2006-07-24</date>
      <author>
        <hparty>
          <id S="ID-HCPARTY" SV="1.0">1234567890</id>
          <cd S="CD-HCPARTY" SV="1.0">persphysiotherapist</cd>
          <firstname>Dirk</firstname>
          <familyname>Masseur</familyname>
        </hparty>
      </author>
      <iscomplete>true</iscomplete>
      <isvalidated>true</isvalidated>
    </transaction>
    <item>
      <id S="ID-KMEHR" SV="1.0">1</id>
      <cd S="CD-ITEM" SV="1.0">encountertype</cd>
      <content>
        <cd S="CD-nomenclature" SV="***">560011</cd>
      </content>
    </item>
  </folder>
</kmehrmessage>

```

Figure 1. An example of a kmehr-bis message for the physiotherapy registry

1.2. Phase 2: Traceability and inalterability

One of the major concerns of the project was that the rationale behind the electronic registry would conform well to the common practice of the physiotherapist. Therefore we opted for a periodic recording of a standardized “image” of the controllable data, in the patient database of the software-system, into the kmehr registry messages. This recording is done in analogy with accountancy systems where the same method is used at the moment of the production of an invoice. This recording still leaves room for the possible corrections to the physiotherapists act within a practical time frame of 12 weeks after the act.

These periodic recordings consist of all the acts, the physiotherapist executed over the past 12 week and are automatically generated every week, starting with the first use of the software package. Besides those automatically generated messages, it is possible to manually create a new recording of the registry at any given moment, which permits the controlling authorities to check whether these messages are created in the correct way.

While in the past, when a paper version was used, changes to the registry were fairly easy to notice. Solving the problem of inalterability of electronic files is not such an easy operation. To overcome this problem we decided to use the SHA-256 hash code as a secure fingerprint of the different files to check if the original message was altered. This hash code which is calculated for each message, each time it is produced, is then stored into a variable, together with a time stamp, within the software package itself. In a later stage this code could be sent automatically to the controlling authorities, to fulfill complete security.

The concept of traceability was fixed by incorporating the SHA-256 hash code of the previously created recording into the new registry file. This procedure is illustrated in figure 2. In this way two time series containing hash-codes appear, one as a variable within the software package, the other in the registry files itself. These two series have to be the same and can at every given moment be tested by manually creating a SHA-256 hash code using freely available software. This procedure makes it possible to trace modifications:

- Justified modifications to the saved data concerning acts of the physiotherapist by making use of the content of the messages
- Possible fraud within a registry message via inconsistencies between the time series of the SHA-codes within the messages on one hand and the variable containing the saved (or in the future transmitted) SHA-codes on the other hand.

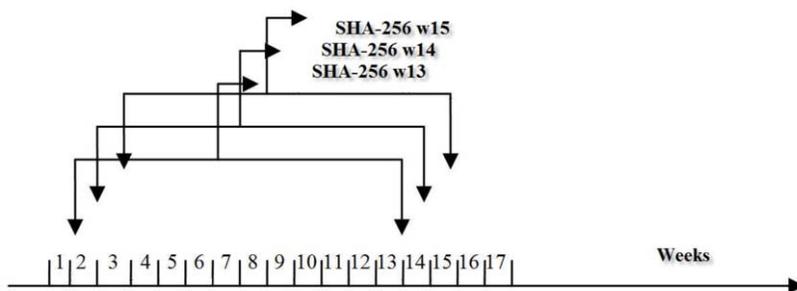


Figure 2. Use of SHA-codes in automatically generated registry messages

1.3. Phase 3: test criteria and control mechanism

The implementation of this procedure was realized by the formulation of test criteria that were used within the certification procedure [7,8] for software packages for the management of physiotherapy records, issued by the “Belgian Ministry of Health, Food Chain Safety and Environment”, organized in November 2006. The following test criteria were used:

- Registry messages must be generated weekly on an automatic basis, but can be generated manually whenever suitable
- Registry messages must be stored separately for each physiotherapist using the software package
- The variable that contains the SHA-256 code of the last generated registry message, is stored for each physiotherapist and a timestamp is added
- Starting from the registry messages, the software package must be able to generate a list (show on the screen and print) of all acts of a specific day sorted by patient name or all acts for all patients sorted by date for any give time period.

The software packages were evaluated on the basis of (amongst others) these criteria by making use of a test scenario that consisted of a number of fictitious patients that had to be entered into a blanc system. At certain intervals a registry message had to be generated. Making use of a specially developed XSLT schema, these messages were tested on consistency of content as well as for accuracy of the SHA-256.

2. Discussion

With the KINELECTRICS project we developed a system that, in a fairly easy way, made the transition from a paper based to an electronic registry for physiotherapists possible. Special attention was given to the user-friendliness of the method, making sure that physiotherapist that are using any of the eighteen certified software packages will not have to worry about producing their monthly registry, this works automatically. This corresponds with almost 5000 physiotherapists (30% of total) who are generating the registry messages automatically. On the other hand, this initiative can be an

incentive for the non-users to start using a certified software package. Or else they must make the effort to comply with the digital registration requirements themselves.

The end result is a clearly outlined method for generating an electronic registry, that meets all the legal constraints, as well as enabling to control the traceability and the inalterability of the produced documents. At this moment the controlling authorities are able to check the integrity of the generated registry messages, possibly together with the history of the SHA-variable. Besides that the software systems are able to present a list of all acts per date and a list of the acts of one patient for a certain period, with standard XML tools. A history of all legitimate adaptations within a certain period can also be produced. The methodology was thoroughly checked within a test environment and during the homologation sessions. The next step is to perform an in-the-field analysis of the produced messages and compare the results with the data found in the physiotherapist's records.

The provided XSLT schema together with the freely available source-code of the SHA-256 hashing procedure form a basis that can allow the controlling authorities to create additional programs that will allow them to track if each the SHA-256 code in each produced message corresponds with the hash code of the previous message. This will also allow them to put together a list of SHA-256 codes originating from the registry messages and in a simple way check their correspondence to the history of the SHA-variable within the software package.

References

- [1] Pekka Ruotsalainen, Bryan Manning: A notary archive model for secure preservation and distribution of electrically signed patient documents, *International journal of medical informatics* 76 (2007), 449–453
- [2] Marcel Lucas Müller, Frank Ückert, Thomas Bürkle, Hans-Ulrich Prokosch: Cross-institutional data exchange using clinical document architecture (CDA), *International journal of medical informatics* 74 (2005), 245-256
- [3] Katehakis DG, Sfakianakis S, Tsiknakis M, Orphanoudakis SC: An Infrastructure for Integrated Electronic Health Record Services: The Role of XML (Extensible Markup Language), *J Med Internet Res* 2001;3(1):e7 <http://www.jmir.org/2001/1/e7/>
- [4] Anne-Marie Rassinoux, Christian Lovis, Robert Baud and Antoine Geissbuhler: XML as standard for communicating in a document-based electronic patient record: a 3 years experiment, *International Journal of Medical Informatics* 70(2-3) 2003, 109-115
- [5] Frédéric Achard, Guy Vaysseix, Emmanuel Barillot: XML, Bioinformatics and data Integration, *Bioinformatics* 17 (2) 2001, 115-125
- [6] Kind Messages for Electronic Healthcare Records Belgian Implementation Standard <http://www.health.fgov.be/telematics/kmehr/>
- [7] Ronald Buyl, Anouk de Smedt, Marc Nyssen: Quality labelling of Electronic Medical and Paramedical Record Systems: A Current Status, *J. Qual. Life Res*, 3 (3)2005 78-81
- [8] Ronald Buyl, Anouk de Smedt, Marc Nyssen: Evidence based testing of electronic medical record systems for general practitioners, *Family Medicine and Primary Care Review*, 8 (2) 2006
- [9] Izquierdo J. N., Schoenbush V.J., The potential and limitations of data from Population-Based State Cancer Registries, *Am. J. Public Health*. 90 (2000), 695-698

Computer Support for Shared Care of Diabetes: Findings from a Danish Case

Keld BØDKER¹, Maren Fich GRANLIEN
Computer Science, Roskilde University, Denmark

Abstract. Shared care has been announced as an effective model for care of chronic diseases. In the paper we discuss various facets of IT support for shared care of diabetes. Based on an empirical study of a project in Denmark we identify various challenges involved with IT support of shared care; structural problems and lack of attention to general practitioners' practice and to clarify the meaning of *shared care* in the actual project. We propose the importance of reaching agreement as to what is shared and we suggest a distinction between two levels of shared care: an *epidemiological level* where hospital specialists and national authorities need to monitor the disease based on data from the general practice; and a *operational, daily level* where the professionals dealing with an individual patient and the patient share information to help the patient deal with his/her situation on a everyday basis.

Keywords. Disease mgt, Communication, Strategic plans, shared care, diabetes

Introduction

Chronic diseases are affecting a growing numbers of patients, and dealing with these diseases is imposing growing costs to the healthcare budgets in western countries. Despite the substantial resources allocated to healthcare, there is a general and growing dissatisfaction with the apparent lack of coordination, continuity and support in the care of individual patients. This is particularly the case when treatment involves several healthcare providers, i.e. hospital departments, outpatient clinics, homecare providers and GPs, which is the case with chronic diseases. On this background *shared care* has been introduced as a concept and general idea to improve the quality and efficiency of complex care services involving the combined efforts of a variety of healthcare agencies and professionals. Shared care can be defined in the following way: "Shared care applies when the responsibility for the healthcare of the patient is shared between individuals or teams who are part of separate organizations, or where substantial organizational boundaries exist" [1]. Other authors define the concept differently, but we find this definition to capture the overall ideas quite well; however it is also clear that the definition is quite vague and needs to be operationalized for specific programs.

¹ Corresponding author: Keld Bødker, Roskilde University, bldg. 42.2, Universitetsvej 1, DK-4000 Roskilde, Denmark. E-mail: keldb@ruc.dk

Shared care programs focus on improving coordination, collaboration and knowledge sharing among healthcare professionals as well as involving patients as active participants in these processes. Research has shown that shared care programs can improve the quality and efficiency of care significantly [2], but also that the implementation of such programs is a long-term and difficult process [3-5]. In Denmark, IT-supported shared care is a central part of the national IT strategy and the action plan with regards to the national initiatives concerning diabetes [4].

The aim of the paper is to discuss the concept of shared care in relation to dealing with diabetes treatment, and in particular to illustrate the challenges involved in designing computer support for shared care. We draw upon an empirical study of a Danish project to discuss some of the important challenges involved in such projects.

1. Case Study and Methods

A case study of the implementation of SharedDiabetes² is the empirical basis for our discussion of challenges involved in providing computer support for shared care in diabetes treatment. The system was developed and implemented in the central part of one of the five Danish regions and is now being spread to the rest of the region - from now on referred to as Region A. The system is being promoted as a system for supporting shared care of diabetes, i.e. containing data and functionality relevant for the treatment of patients with diabetes. The system is a web-based application working independently of the other systems of the care providers with only little exchange of data. The system is integrated with the laboratory system used in the region's hospitals, while integration with the GPs' electronic patient record systems is based on EDIFACT messages to the GPs' systems. At the time of our investigation the outpatient clinics did not have an electronic patient record and used SharedDiabetes as their main system for diabetic patients together with the hospital patient administration and medication system. SharedDiabetes is used by all outpatient clinics, a substantial part of the GPs, and has recently been made accessible to patients through the Danish National Public Healthcare Portal (sundhed.dk).

SharedDiabetes was developed in response to an action plan for diabetes from the Danish National Board of Health in 1994 [6]. Following this plan, a regional diabetes committee was established, diabetes outpatient clinics were restructured to reflect a team oriented care, and a need for an IT system that could help monitor and support the treatment as well as provide a means for quality assurance was identified. The regional diabetes committee established a project with a large pharmaceutical company that had a system under development. Later the development of the system passed to a large IT company. The project involved representatives from the various stakeholder groups, e.g. diabetes doctors and nurses from the outpatient clinics and GPs, mainly two diabetes practice consultants i.e. GPs who - beside their consultation - work with improving the collaboration between diabetes outpatient clinics and general practice. To understand the many stakeholders it is necessary to understand the treatment of diabetes in Denmark.

When a suspicion for diabetes with a patient is discovered - at a hospital or in the GP's practice - the patient is referred to the diabetes outpatient clinic to have the diabetes diagnosed properly. Here the patient upon diagnosis receives initial treatment

² a pseudonym.

with the goal of becoming *well regulated*, i.e. achieve a stable situation by a combination of medication, a changed diet and physical exercise. In the outpatient clinic the patient is treated by a multi-disciplinary group of specialists: diabetes doctors and nurses, dieticians, ophthalmologists (eye specialists), and podiatrists (feet specialists).

When the patient is well regulated, typically within less than a year, the patient is discharged to his/her GP. This is the workflow for type 2 diabetes - by far the most widespread (type 1 patients are treated in the hospitals). Upon discharge, type 2 patients are seen in the GP on a regular basis, every three months and a more thorough control every 12 months. Once a year the GP refers the patient to see an eye specialist and a foot clinic. The GP evaluates indications from these specialists as part of the annual status. Late complications of diabetes are serious to the individual patient and costly to deal with for the healthcare system. Effective care and the patient's active participation are central to reduce the late complications, and thereby improve the quality of life for the patients and lower the costs for society.

The empirical investigation took place throughout 2006 and comprised a series of interviews, observations and document analyses. Interviews were semi-structured and lasted 45-75 minutes; they were recorded and later transcribed. In region A, three GPs were interviewed about shared care (referred to as GP1, GP2, GP3). In addition to the interview, we conducted a think-aloud test with the GPs where they went through the system on the basis of a short consultation scenario. Further, two persons from the IT company and one from the regional committee were interviewed. Observations at an outpatient clinic were carried out over three days - in total approximately 15 hours of observation was documented by field notes. Interview transcripts and field notes from the observations were analyzed using a coding process inspired by the theoretic sampling technique known from grounded theory [7].

2. Results and Discussion

The project in Region A was well integrated with the regional activities of diabetes treatment, as evidenced by [6] and our interviews; the system is actually quite well designed and technically sound; and a leading medical doctor from the regional University Hospital has recently claimed remarkable results in terms of significantly less late complications to the shared care initiatives (TV2 News, May 15, 2007). However, findings from the study also point to a number of challenges for IT supported shared care, especially related to GPs' use of SharedDiabetes. Some of these challenges are related to structural issues, some are related to the insufficient involvement of key actors and some are related to our central questions of what is shared or what should be shared.

2.1. Structural problems and lack of attention to general practitioners' practice

A large part of the structural problems comes down to the fact that GPs are private businesses while diabetes outpatient clinics are part of public hospitals. In Denmark virtually all GPs have electronic patient records, and there are more than 20 different EPR systems in use in GP's clinics. Currently the Danish regions owning the public hospitals are developing and implementing full-scale hospital EPR systems, but when the SharedDiabetes project started, and when the system was first implemented in the

region, there was no EPR system at the region's hospitals. For that reason SharedDiabetes was designed for use as the main system for diabetes patients, in the outpatient clinics as well as in the GPs.

The GPs, however, have their own EPR system as their main system. This means that the use of SharedDiabetes in GP's clinics is on top of their EPR system involving extra work and overhead. When seeing a diabetes patient the GP has to log into SharedDiabetes by opening a browser window, check data from previous consultations, register the relevant data from the current consultation, and then return to the EPR system. This extra work was actually acknowledged. As part of the project the services to be performed by GPs were regulated and included in the financial agreements between the GPs and the region. However, in the daily use the lacking integration causes frustrations, as illustrated by the following quote. "I was quite insisting on this matter, but it was technically not possible to do this at this point in time. I am of the opinion that a [shared care system] should have been developed where the application should be within our computers" (GP3). As noted by the GP, achieving a tight, or *seamless*, integration between SharedDiabetes and his EPR system requires a vast amount of resources. Given the software architecture decisions in the project this would involve all the 20+ providers of EPR systems for GPs.

In general the GPs find that SharedDiabetes does not support their way of working. The system is designed for use *during* the consultation to facilitate involvement of the patient (to help the patient become active in his/her own treatment). This is how we observed the system being used at the outpatient clinics. However, the GPs we interviewed never used the system during a consultation but preferred to consult the system before and register the data after the consultation. "I would rather use the time – and I might sound a little self-righteous now – but I prefer using my time with the patient and then use the computer either before or after [the consultation]. That's what I do generally" (GP2).

Due to the fact that the information input to SharedDiabetes is not integrated with the GPs' electronic records, and the fine level of granularity of data requested by the doctors at the outpatient clinics, using the system becomes tedious extra work in the GP clinic. This is reinforced by the relatively low frequency of diabetes patients in the general practice. Less than 5% of a GP's consultations regard diabetes [8], which means that s/he typically has 50 patients or less with diabetes and only one fourth of them is also being treated by the outpatient clinic. What this means is that using the system never becomes routine, the GP always has to use extra attention - for example to locate a specific field.

These findings resemble experience from other studies of the uptake of clinical information systems, for example [9]. The implication is that by not attending to the GPs' professional practice regarding use of computers in their consultations, risks are high that GPs will only use a shared care system to an extent fulfilling the minimal requirements in agreements between the region and GPs. From this, one could hypothesize that a decreasing number of GPs would use the system as intended. We have not been able to test this hypothesis, as we have only been able to obtain data on the number of GPs having followed a course on SharedDiabetes and hence obtained a password to the system (more than 50% of GP clinics [6]).

Another implication from the findings could be formulated as a question: Could we approach some of the challenges head-on? If, for example SharedDiabetes was extended to include other chronic diseases, how would that affect the uptake in GP? We can of course only speculate in trying to answer such questions. It is clear that this

would address some of the challenges: the GPs would use the system more often and use would thus become more routine, and it would become more feasible to address the question of proper integration with the GPs' EPR systems.

2.2. What does shared care actually mean?

We have illustrated a tension between requirements from the outpatient clinic care providers towards a fine granularity of data in SharedDiabetes and the GPs' wish for simplicity and ease of use. From this, we could easily foresee a situation where diabetes specialists blame GPs for not using the system as intended, while GPs think that the system - with a level of detail of information and mode of interaction based on the outpatient clinics' situation - illustrates the privilege of resources needed to attend to details of care in public hospitals without attending to the practicalities of running a GP clinic. We suggest the tension might be more productively used to understand the conditions and challenges for developing and using shared care applications.

From our analysis we suggest the key is not only to understand *what is shared and with whom*, but also to reach an agreement on *what kind of shared care* we want to support. From the perspective of dealing with the individual patient as a diabetic, care may be shared between the primary actors involved, i.e. the GP, the various specialists in the primary sector and home care providers. However from the perspective of the GP it only rarely makes sense to talk about care as shared with the outpatient clinic. This means that the working interpretation of shared care applied in the project, "we talk about shared care when a diabetic has received treatment, or has been registered, in both the outpatient clinic and in the GP within the last three years [6, our translation]", is not anchored across all shareholders, and potentially does not represent all shareholders' view.

From a specialist perspective many data are relevant and valid. The data are needed to support a long-term level or an *epidemiological level* of shared care where hospital specialists and national authorities need to monitor the disease based on data from the primary sector. However to the GPs, the low frequency of diabetic patients and the mode of interaction with the system come at odds with their professional practice. In our interpretation, they acknowledge a need for support on an *operational, daily level* where the professionals and the individual patient share information to help the patient deal with his/her situation on an everyday basis. Many GPs state that the system does not contribute any new, and to them, *relevant* information. It is actually hard to find support for shared care in the treatment of diabetes as implemented in SharedDiabetes among the GPs in the study. Their understanding is that only seldom the healthcare professionals share their care across sectors; instead they see the patient as either being treated at the outpatient clinic or at the GP's clinic *for the time being*. The diabetic patient, so to say, does not constantly move back and forth between the GP and the outpatient clinic.

From this perspective, it would make much more sense to devise IT support for *exchanging data* than to talk about shared care in the strict sense seen from the GP when talking about providing data for an epidemiological purpose. If the system is to support the GPs professional practice of providing care in collaboration with the patients themselves, then what is shared might differ quite significantly.

3. Conclusion

It is clear that a more efficient treatment of diabetes is required due to the rising number of diabetics and accompanied risks of late complications. The case described is a welcomed initiative in that direction. The system is technically sound and well designed, and the project demonstrates that many relevant and necessary steps have been taken. However, from the GP's perspective SharedDiabetes supports the regional outpatient clinics' monitoring of diabetes and provide data for proactive actions towards late complications. The system does not support the GPs' professional practice of dealing with diabetics due to a number of reasons: lack of need for shared care in terms of shared data, lack of attention to GPs medical practice, and lack of proper integration with the GPs' EPR systems.

Our analysis of the case study identified multiple levels of shared care: the national/regional level with a mainly epidemiological purpose and a local level centered around care of the individual diabetic patient. Shared care in the sense of providing care for the individual diabetic by a group of professional actors with different institutional belongings is certainly feasible but would require a different set up from the one described with SharedDiabetes.

Acknowledgments

We would like to thank Sine Marie Øelund and Louise Baunsgaard Koch who together with the second author, Maren Fich Granlien, have conducted the empirical study that formed the empirical basis for this paper.

References

- [1] Pritchard, P. and J. Hughes, *Shared care - the future imperative*. 1995, London: Royal Society of Medicine Press.
- [2] Rubak, S.L., J. Mainz, and J.M. Rubak, [*Shared care--integrated cooperation across the sector boundaries about patient's course. Modern health care system cooperates*]. *Ugeskr Laeger*, 2002 **164**(45): p. 5256-61.
- [3] Bansler, J., E. Havn, and F. Kensing, *IT-support for shared care*, in Hejlesen, O. and C. Nøhr (Eds): SHI2005 Proceedings *3rd Scandinavian Conference on Health Informatics*. 2005: Aalborg: p. 9.
- [4] Granlien, M.F. and J. Simonsen, *Challenges for IT-supported shared care: a qualitative analyses of two shared care initiatives for diabetes treatment in Denmark*. *Int J Integr Care*, 2007 **7**: p. 1-13.
- [5] Mur-Veeman, I., I. Eijkelberg, and C. Spreeuwenberg, *How to manage the implementation of shared care: a discussion of the role of power, culture and structure in the development of shared care arrangements*. *J Manag Med*, 2001 **15**(2): p. 142-55.
- [6] Bassett, B., et al., *Diabetesbehandling - den fynske model*. *Månedsskrift for praktisk lægegerning*, 2006 **84**(1): p. 33-46.
- [7] Glaser, B. and A. Strauss, *The discovery of grounded theory: strategies for qualitative research*. 1967, New York: Aldine de Gruyter.
- [8] Vedsted, P., et al., *Almen lægepraksis i Danmark [General Practice in Denmark]*. 2005, Århus: Tidsskrift for praktisk lægegerning og forskningsenheden for almen praksis.
- [9] Short, D., M. Frischer, and J. Bashford, *Barriers to the adoption of computerised decision support systems in general practice consultations: a qualitative study of GPs' perspectives*. *International Journal of Medical Informatics*, 2004 **73**(4): p. 357-362.

From Documents on Paper to Electronic Medical Records

Lino CARRAJO ^a, Ángel PENAS ^a, Rubén MELCÓN ^a, Fco. Javier GONZÁLEZ ^a,
Eduardo COUTO ^a

^a *R&D&I, Information Technology Department, CHU Juan Canalejo, A Coruña, Spain*

Abstract. This paper describes the creation process of an electronic medical records (EMR) application in the Juan Canalejo University Hospital Complex (CHUJC). From the knowledge acquired through the observation of the traditional processes of managing the Patients medical records on paper a tool was developed which in principle was thought of to classify electronic documents associated to a patient and to which different functions of medical work have been subsequently added: visualizing clinical documents of patients, creation of new documents and following the development of patients.

Keywords. Health professional workstation, Archival-repository systems for medical records- EPR-CPR-EMR, Knowledge-based systems, HIS management.

Introduction

People working in a hospital need to have access to clinical information of those patients who are under their responsibility. This information must be shown in a structured way so that work on it becomes easier, allowing a quick access to documents needed for a diagnosis, treatment and monitoring of patients.

In 2001, in the CHUJC clinical information was kept on paper, Word documents stored in the network directories and DBASE IV applications. All those resources had no link to relate clinical documents to the patient. The means used by health workers to diagnose a patient and to follow the development of illnesses and diseases was paper.

The need for the integration of the different documents and clinical reports of a patient under the same control has arisen. It is necessary to offer health workers a way of electronic access to patients' records which previously only existed on a digital format, without being necessary to open the large number of departmental applications and searching for the data that the patient had in any of them.

1. Materials and Methods

1.1. Objectives

The objective was to develop software which would become the only portal for the hospital's staff to obtain information about patients' health. Information has been

classified according to different criteria, allowing the staff to have access to it in a faster and more intuitive way than if they were visualising it on paper or through different applications.

On the other hand our intention was to facilitate the creation of new medical reports under a homogeneous format for the whole organization: the use of a single corporate tool, the classification of reports according to types, the production of useful types of reports which will represent all the medical information of the patient, the same templates for identical types of medical reports.

It was necessary to develop a tool for the hospital's staff which could be adapted to the manner in which they work and that would make it easier tracking the diseases and illnesses of the patients and the communication among the different specialists, in short, keeping a well-structured of Electronic Medical Records.

The information to be managed by the new system of information is confidential, classified as High-level, the maximum level, by the Spanish Royal Decree 994/1999, 11th June on Security Measures for Computer Files Containing Personal Data. Such regulation is in force as well the LOPDCP (Organic Law 15/1999, on the Protection of Personal Data), and according to which access to personal data is regulated to guarantee the right to privacy which every person is entitled to. The system was designed to meet the requirements specified by those regulations and to establish two basic objectives:

1. Users should only have access to the information necessary for carrying out their work. In case they required access to another kind of information, their reasons for doing so should be justified.
2. Every time they have access to the personal data of the patients, this must be recorded. This register must be accessible to those people who are responsible for the security of the information services.

1.2. Technical Design of the Solution

1.2.1. Architecture of the System

To obtain a system that would fulfill the policy of software development in our organization and that at the same time was strong, had a high level of performance and was able to endure a high number of users having access at the same time, we used a multilayered architecture based on JEE technology:

- Data were stored in a relational database capable of storing binary documents and enduring a high transactional charge.
- The layer of access to data and the logic of business were implemented using EJB's technology according to the requirement of JEE 1.4.
- The layer of presentation was implemented by using standalone client technology. The platform on which the client should run was Microsoft Windows, since one of the requirements of the system was that the production of new medical reports had to be made by using the tool Microsoft Word™.

The first problem which had to be solved when using a of non-web client's architecture was the installation of the software in every workstation and the distribution of the revisions that would be made with the evolution of the product. Two solutions were adopted:

1. The installation of the client application in Citrix servers.
2. The use of JNLP (Java Network Launching Protocol) for the distribution of the revisions. The first task done by the application is to connect itself to a web server by using this protocol and checking the existence of changes in the software packages. That being the case, the new packages are downloaded to the local system and normal execution continues.

For the communication of clients (presentation layer) with the “Business Layer” protocol RMI/IIOP is used. This protocol transmits data in binary system between client and server, which allows a more efficient communications compared to a solution based on Web Services.

The organisation has at their disposal a repository of users with access to those information systems registered in Microsoft Active Directory. LDAP (Lightweight Directory Access Protocol) is used to delegate the authentication of the users in the system and the policy of use of passwords in the organization.

In order to increase the performance of the system, especially the communications between client and the application’s server and in the access to data, the following patterns of design were implemented:

- Data Access Object (DAO), for the access to databases.
- Data Transfer Object (DTO), for the encapsulation and transmission of data among the different layers of the system.
- Session Facade, for establishing the remote interfaces which are available and isolating the clients from the implementation of access to data.
- Service Locator, for efficiently locating from the clients the remote resources of the business logic.

1.2.2. Tools Used

A great number of tools have been used for the development and production of the solution:

- Borland JBuilder v.6-v.2006, as integrated development environment (IDE).
- Informix Dynamic Server v9.40 as main Database Management System (DBMS) for storing the information generated by the information system. The new system should obtain information from other DBMSs which were storing the patient’s medical information: Microsoft SQL Server, SyBase.
- IBM Websphere Application Server v5.1 and v6.1, as applications’ server J2EE for the display of the business objects (EJB’s) and access to data.
- Sun JDK 6.0 for the production of the user’s graphic interface through the use of Swing library.
- JFree Chart, for the production of evolutionary charts.
- Jakarta FOP, for the production of PDF reports from XML documents and XSL style sheets.
- Jacozoom, for access through JNI (Java Native Interface) to native elements of the Windows platform. Its main use is to operate the tool Microsoft Word from a Java application.

2. Results

In 2001 the design of the application was started and the application began to be produced at the end of 2004. The application was called “*Gestión Documental*” (Document Management) since its function at the beginning was precisely that, to manage the documents of the patients. After 12 revisions and 3 more years of development the application became the medical workstation, electronic medical history tool used by the professionals of the Juan Canalejo Hospital.

The main functions achieved by the application so far have been:

- Doorway of access to administrative data of patients, their medical histories and reports of tests managed by the system or generated by other department systems, such as cardiology, urology, radiology, pathological anatomy, medical tests laboratories, digestive tests, nursing cares, digital medical image... The system does not store the information from other subsystems. When it is required by a client, the system has access to every subsystem to verify the existence of the information required, and that being the case, it can recover it. The information of the patients is not duplicated. Every subsystem stores and keeps the information which it manages.

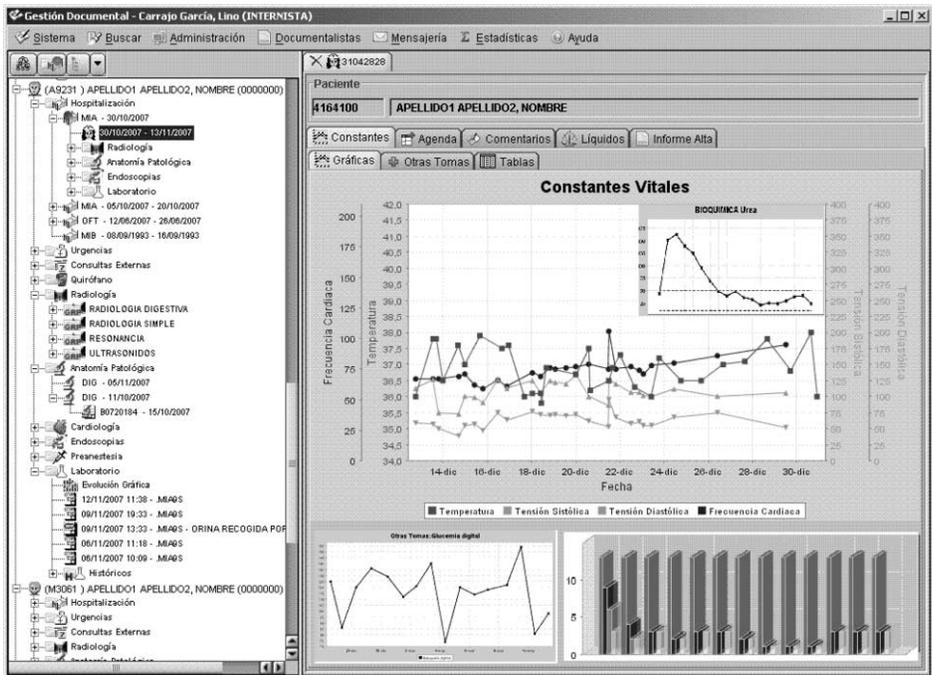


Figure 1. Structure of Clinical History and graphical evolutions

- It supplies a tool for the production of new medical documents. The system allows to produce documents and associate them to the administrative medical case to which are related. To create reports a high integration with Microsoft Word™ (2000, Xp, 2003 and 2007 versions) has been undertaken, since this is the corporate text processor. The system operates the management of the documents: production, edition, closing, printing... Every operation done in relation to a document is managed by the system. Moreover there are other functions:
 - o The use of documents' templates stored in the system and classified according to type of report, speciality and, in some cases, pathology. The system manages an unlimited number of templates.
 - o Automatically-incorporated data into reports utilities have been implemented: the patients' administrative data and medical cases data, results from medical tests in laboratories, conclusions from radiology or pathological anatomy tests.
 - o Signature of the document: basic electronic signature (log in / password) and / or acknowledged advanced electronic signature (digital certificate issued by the National Mint- FNMT).
 - o Documents' versioning utilities: addenda, new version and new replication.
 - o Coding of the reports according to the diagnosis and the procedure. The international system CIE-9 is used. This system is used to search for reports according to the pathology.
- Production of graphic elements to help with the diagnosis of the patient and the management of the surgery waiting and medical external consultations list registry.
- Statistical data operating from the production of reports.
- A system of notifications has been incorporated so that the different professionals who use the application can exchange messages. The notifications can be generated manually by the users or automatically by the system to notify events to the users: system stoppages, new versions of documents previously consulted by the user...
- Medical interconsults system of management for hospitalization services. If a professional needs to consult another specialist, he or she could do it through the application: a circuit of demand, reception, reply and tracking of the consultation has been established.
- Module of Subscriptions to medical alerts. Doctors do no longer have to check when new medical information about a patient is available. After subscribing, as soon as new information is available, the system automatically sends a message indicating the new kind of information available. There are three different ways of notification: through the application, by using emails or through a SMS to a mobile phone.
- The system has been provided with a mechanism for controlling errors. When an error is made in any module of the system, an email is automatically sent to the IT department with the details of the error.

Table 1 shows the degree of penetration of the tool in the day-to-day medical practice at the Juan Canalejo Hospital:

Functionality	Number of items
Well-defined users within the system	2,800
Users who regularly use the system	2,200
Users that have daily acceded to the system	965
% implementation within the organization (consultation mode)	99%
% introduction within the organization (production of documents)	80%
System's concurrent users (maximum)	400
Medical histories uploaded	6,600 – 7,000 / day aprox. 150,000 / month
Reports produced through the tool	274 / day 6,000 – 6,500 / month
Reports from laboratory medical tests consulted	8,300 – 8,700 / day
Reports to Pathological Anatomy accessed	525 /day
Reports to Radiology accessed	600 /day
Studies of digital medical image accessed	900 – 1,100 / day
Cases of nursing accessed	600 - 700 / day

Table 1. Use of the tool (average data between 01-Feb-2008 and 28-Feb-2008)

3. Conclusion

The system which has been created provides the health staff in our organization with a global and structured view of the medical history of the patients, allowing an agile and real-time access to the information, avoiding the delays caused if the documentation had to move around on paper or, in some cases, if it went astray or was damaged.

Finally, we would like to indicate that the system continues developing and incorporating new revisions periodically (approximately once every two months). The improvements in the systems are not only technical, but they incorporate new functions demanded by the users as well as documents generated (produced) by new departmental systems which become integrated in the organization.

References

- [1] User Manual "*Gestión Documental*". I+D+I CHU Juan Canalejo. A Coruña. SPAIN.
- [2] Administration Manual "*Gestión Documental*". I+D+I CHU Juan Canalejo. A Coruña. SPAIN.
- [3] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995.
- [4] Spanish Royal Decree 994/1999, 11th June, on Security Meas. for Comp. Files Containing Personal Data.
- [5] Spanish General Health Law 14/1986, 25th April.
- [6] Spanish Law of Digital Signature 59/2003, 19th December.
- [7] Galician Regulatory Law of Informed Consent and Clinical Records of Patients 3/2001, 28th May.

An Archetype-based Testing Framework

Rong CHEN^{a,b,1}, Sebastian GARDE^c, Thomas BEALE^c, Mikael NYSTRÖM^a, Daniel KARLSSON^a, Gunnar O. KLEIN^d, Hans ÅHLFELDT^a

^a *Department of Biomedical Engineering, Linköping University, Sweden*

^b *Cambio Healthcare Systems, Sweden*

^c *Ocean Informatics UK, London*

^d *Karolinska Institutet, Sweden*

Abstract. With the introduction of EHR two-level modelling and archetype methodologies pioneered by openEHR and standardized by CEN/ISO, we are one step closer to semantic interoperability and future-proof adaptive healthcare information systems. Along with the opportunities, there are also challenges. Archetypes provide the full semantics of EHR data explicitly to surrounding systems in a platform-independent way, yet it is up to the receiving system to interpret the semantics and process the data accordingly. In this paper we propose a design of an archetype-based platform-independent testing framework for validating implementations of the openEHR archetype formalism as a means of improving quality and interoperability of EHRs.

Keywords. Computerized Medical Record Systems, EPR-CPR-EMR, Standards, Knowledge-based systems, Archetypes, openEHR, EHR, Testing, Semantic Interoperability

Background

The innovation of two-level modelling of Electronic Health Records (EHRs) and archetypes [1] pioneered by openEHR [2] and standardized by CEN/ISO [3] brings us one step further towards semantic interoperability of EHRs [4]. Instead of being hard-coded into proprietary software by software developers, clinical content models are expressed in the Archetype Definition Language (ADL) [5] and authored by the clinical professionals themselves. Archetypes are used at runtime by EHR systems to validate user data entry and query fine-grained data in the EHR. Archetype-based EHR systems are highly adaptive and can evolve when clinical requirements change over time since volatile clinical requirements are captured in archetypes while software systems are built using only the stable openEHR information model and archetype language. Archetypes are expressed in a standardized formal language so they are machine-interpretable and can be shared between systems. This makes the semantics of EHR data available not only to other EHR systems but also to surrounding systems.

¹ Corresponding Author: Rong Chen, Department of Biomedical Engineering, Linköping University, Sweden; E-mail: rong.chen@imt.liu.se

Archetypes can be considered similar to software written in a declarative programming language. They define what should be recorded in an information repository, without defining procedural semantics. Both the validation and initial creation functions can be realized by a software component based on the Reference Model (RM) and Archetype Object Model known in openEHR as a ‘kernel’, using the semantics defined in the archetype formalism. In an archetype-based system, we can consider the archetypes and the kernel together as the entity providing the information semantics of the system. Thus, it is desirable to develop unit tests for each archetype that is developed according to its domain model together with the kernel to use it.

Since a ‘kernel’ is crucial to any archetype-based system, any defects or misinterpretation of the specifications in the implementations could impede archetype-based data processing and data exchange between systems. It is therefore important to validate the kernel implementations to ensure archetype semantics are carried out correctly in different implementations across platforms. Traditional hand-crafted unit test code in its classic form is not useful, since it is implementation- and platform-specific. Additionally, since archetypes are shareable and platform-independent, it seems reasonable to have a platform-independent way of testing archetypes and archetype-processing software. This leads to the following requirements of a testing framework for archetype-based systems:

- 1) Tests are platform-independent and implementation-independent;
- 2) Tests are archetype-based and preferably auto-generated.

This paper presents a platform-independent archetype-based testing framework that fulfils these requirements.

1. Methods

In software engineering, unit testing is a procedure used to validate that individual units of software are working properly. It creates a ‘safety net’ for the software when it evolves throughout its lifespan. We used the design patterns from Beck’s unit testing framework [6] and applied it to archetypes system in a platform-independent way.

2. Results

2.1. Design overview

The core requirement for this design is to find a platform-independent way for testing different kernel implementations of archetype-based systems. The scope of the test should cover the core functions of the kernel component, initially including archetype based data creation, validation and path based querying. The testing framework consists of three logical parts (Figure 1).

- The first part is test fixture representation and generation. A test fixture refers to a fixed state in order to run a test and expect a particular outcome. In the context of testing archetype-based systems, a test fixture is equivalent to a state of a reference model object.

- The second part is test case representation and generation. A test case is a unit of testing logic for a specific test scenario. It includes references to test fixtures, the kernel operation to test and the expected result from invocation.
- A test runner can load test cases and test fixtures, execute the specified kernel operation against a particular kernel implementation with given input values and compare the returned value with the expected result to decide whether the test has passed or not.

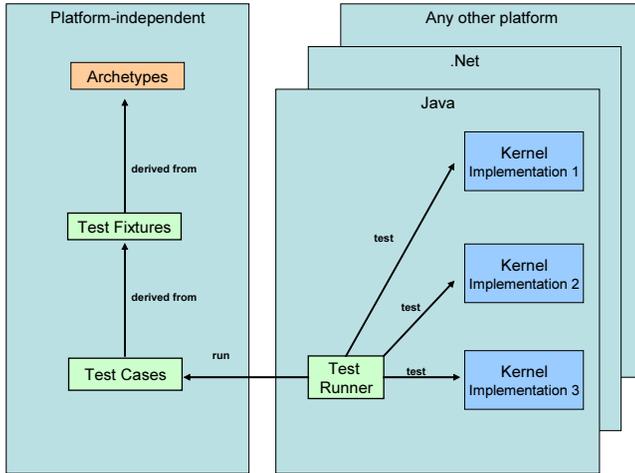


Figure 1. System diagram of archetype-based testing framework

2.2. Test fixtures

One requirement on the design of this testing framework is that it should be platform-independent so that it can be used to validate different openEHR kernel implementations on different platforms. This requires the representation format of test fixtures to be platform-independent. We considered using XML for this, but due to the ambiguities in handling common container data structures like List and Hash, the data Archetype Definition Language (dADL) was used instead. Since dADL is also used to represent archetypes, it is expected to be supported by all archetype-based systems.

One of the difficulties of archetype-based testing is that the number of possible states of RM that are valid according to an archetype can be so large that it is close to impossible to test all combinations. A good strategy of generating test fixtures is needed to have good coverage of valid states of RM and at same time not yield too many fixtures. A reasonable set of test data objects should at least cover typical data that could occur in a production EHR context, as well as some of the likely exceptions.

One design idea is to automatically generate fixtures based on pre-defined rules. Rules can be specific to any type of constraint or object node. Constraints in archetypes can be divided into two categories: 1) object constraints, which are to do with object occurrences and cardinality etc; 2) leaf constraints i.e. constraints on primitive data types such as Integer and String. Test fixtures can be generated by examining constraints on each object node level. For example, one rule can exclude any optional attributes while another can include them. On the leaf constraint level, valid values can be selected by inspecting the constraints. For example, for the integer range “1..10” constraint, a specific rule can include both boundary values and a random value in the

range. The total number of auto-generated fixtures can grow rapidly depending on selected variations and the depth of the object tree. By choosing a good strategy using appropriate rules, it is possible to get a good balance between manageable size of the test fixture and representativeness of the data. It is possible to calculate the total number of fixtures to be generated using a specific set of rules in advance. This gives the possibility to tailor the rules so that a suitable amount of fixtures can be obtained. Another design idea is to combine both manually selected special instances and auto-generated fixtures. This way, test fixtures that are known or suspected to cause errors can be included even if they are excluded by auto-generation rules.

2.3. Test cases

Each test case includes specifications for: 1) the kernel operation to test; 2) the input values to the kernel operation if required; 3) the expected return result from the kernel operation. Both the input values and the expected result can be a reference to a test fixture if a complex object instance is required. In order to test the kernel thoroughly, it is necessary to include not only test cases that expect the kernel operation to return with a result but also those that expect the kernel operation to abort with an error message. Since test fixture generation is based on archetypes, it is reasonable to group test cases around archetypes. It also makes sense to automate the generation of test specifications using existing text fixtures.

Creation of test cases for archetype based data creation can be automated by inspecting test fixtures and computing the input values. For a given reference model object, one test case can be created to test if the same object can be created by the kernel using computed input values and several other test cases can be created with deliberately invalid values to cause a data creation failure from the kernel with anticipated error messages. To generate test cases for data validation, it is possible to use a valid test fixture as is and expect the kernel validation function to return a positive answer. To create a test case that causes a negative result from the kernel, the test fixture needs to be modified to make it invalid. This can be achieved by modifying the object tree in such a way that it violates certain constraints. Path-based query test cases can be generated by extracting all valid paths from a given fixture and then querying the fixture using a reference kernel implementation to obtain the correct values used as the expected results in the test case.

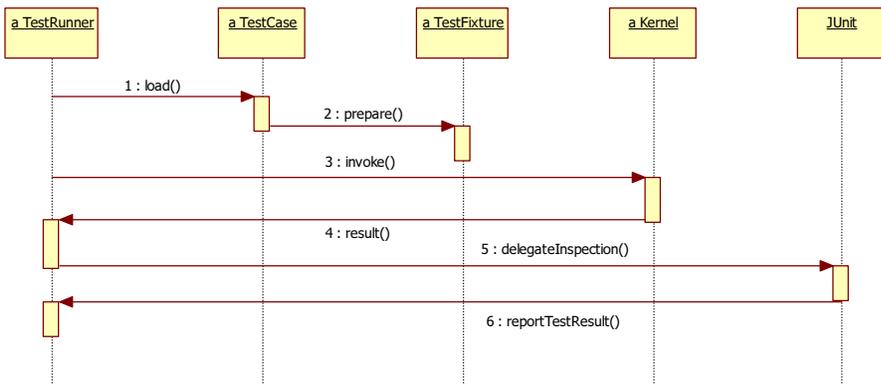


Figure 2. UML Sequence Diagram of a test execution

2.4. Test runners

A test runner can load test fixtures and use test cases as specifications to carry out actual tests against an individual kernel implementation. A typical use case is as follows: a test runner takes a test case and loads according test fixtures with specified input data and expected result. Then it invokes the specified kernel operation according to the test case, takes the output from the kernel and compares it with the specified result to decide whether the test passes or not (Figure 2). The result inspection part can usually be delegated to the underlying unit testing framework, e.g. JUnit². The parsing of test fixtures is done by a dADL parser, which usually exists for an openEHR system. Because a test runner needs to interact with a kernel component through direct method invocations, it needs to be platform specific. But if a kernel implementation can be accessed in a platform-independent way through for example Web Services, even this part of the testing framework can be made platform-independent.

2.5. Implementation

We verified this design by prototyping it using available open source components from the openEHR Java Reference Implementation project [7]. Test fixtures and test cases (Listing 1) are initially manually created based on the blood pressure archetype³. We plan to add support for archetype-based automated generation of test fixtures and test cases later on. A Java test runner will be made available via the openEHR Java project⁴.

Listing 1. Test case specification in XML for testing path based query

```
<PathQueryTestCase>
  <DataInstance>blood_pressure-001.dadl</DataInstance>
  <Test>
    <Path>/observation/data/events[1]/data/items[1]/value/magnitude</Path>
    <Expected>120.0</Expected>
  </Test>
  <Test>
    <Path>/observation/data/events[1]/data/items[2]/value/magnitude</Path>
    <Expected>90.0</Expected>
  </Test>
</PathQueryTestCase>
```

3. Discussion

Compared with traditional testing frameworks, the key innovation of this design is that the test fixtures and test cases are archetype-based and therefore independent of any particular archetype kernel implementation on any computing platform. Because archetypes are machine-interpretable, it is possible to automate the generation of test fixtures and test cases based on pre-defined rules.

One important aspect of system interoperability is to reduce ambiguity of data semantics. The archetype formalism has delivered on providing the semantics to EHR

² <http://junit.org>

³ http://www.openehr.org/svn/knowledge/archetypes/dev/adl/openehr/ehr/entry/observation/openEHR-EHR-OBSERVATION.blood_pressure.v1.adl

⁴ <http://www.openehr.org/projects/java.html>

and surrounding systems. But it is still up to actual implementations to interpret the archetype semantics and apply them during EHR processing. Differences between different archetype formalism implementations pose a potential threat to EHR interoperability. The archetype-based testing framework we propose in this paper can be used for validating different archetype implementations to help ensure interoperability between different systems and platforms.

With the introduction of EHR two-level modelling and archetype methodology, we are entering a new era of developing health information systems. Such systems are more open, adaptive and collaborative than ever before. They are based on standardized information models, shared clinical content models and open sourced core components. This will not only increase the total productivity in the field of health information systems, but also enable us to develop more advanced clinical systems using the semantics intended by the original clinical content models. We believe that EHRs will be central to this new paradigm and their semantics, expressed as archetypes and their interpreters, the kernels, will be of vital importance to this. The correctness of the archetype implementation in different archetype-based systems is crucial to the quality and interoperability of EHRs. An archetype-based platform-independent testing framework is a first attempt to address this issue.

The testing framework is extendable. When the kernel functionality evolves, support for new kernel operations can be added. This can be achieved by adding new operation types in the test case expression format and mapping it to the new kernel function invocations within the test runner. As a next step, we will explore the selection of the best strategy for archetype-based generation of test fixtures and test cases. It is desirable to have an open source implementation of a test fixture generation tool, a test case authoring tool and one test runner for each major computing platform.

4. Conclusion

The platform-independent archetype-based testing framework we propose here is potentially useful for validating different archetype formalism implementations as it can ensure that the semantics of archetypes are interpreted correctly and utilized faithfully across different systems and platforms. Thus this work contributes to the aim of semantic interoperability of EHRs.

Acknowledgements

This work was kindly supported by Cambio Healthcare Systems, Sweden.

References

- [1] Beale T, Heard S. Archetype Principles in The openEHR Foundation Release 0.95. In: Beale T and Heard S, eds. The openEHR foundation, 2005
- [2] Kalra, D., T. Beale, et al. (2005). The openEHR Foundation. *Stud Health Technol Inform* 115: 153-73.
- [3] EN 13606-1:2007 Health informatics - Electronic health record communication - Part 1: Reference model, CEN - European Committee for Standardization. Also published as ISO 13606-1:2008, ISO-International Organization for Standardization.
- [4] Garde, S., P. Knaup, et al. (2007). Towards semantic interoperability for electronic health records. *Methods Inf Med* 46(3): 332-43.
- [5] Beale T, Heard S. Archetype Object Model. (2006), The openEHR Foundation.
- [6] Beck, K. Simple Smalltalk Testing: With Patterns, <http://www.xprogramming.com/testfram.htm>, [cited 2007 November 12]
- [7] Chen, R. and G. Klein (2007). The openEHR Java reference implementation project. In *Proceedings of Medinfo 2007*, *Stud Health Technol Inform* 129: 58-62.

Supervised Approach to Recognize Question Type in a QA System for Health

Sarah CRUCHET, Arnaud GAUDINAT, Célia BOYER

Health On the Net Foundation, Geneva, Switzerland

Abstract. Many attempts have been made in the QA domain but no system applicable to the field of health is currently available on the Internet. This paper describes a bilingual French/English question answering system adapted to the health domain and more particularly the detection of the question's model. Indeed, the Question Analyzer module for identifying the question's model has a greater effect on the rest of the QA system. Our original hypothesis for the QA is that a question can be defined by two criteria: type of answer expected and medical type. These two must appear in the step of detection of the model in order to better define the type of question and thus, the corresponding answer. For this, questions were searched on the Internet and then given to experts in order to obtain classifications according to criteria such as type of question and type of medical context as mentioned above. In addition, tests of supervised and non-supervised classification were made to determine features of questions. The result of this first step was that algorithms of classification were chosen. The results obtained showed that categorizers giving the best results were the SVM. Currently, for a set of 100 questions, 84 are well categorized in English and 68 in French according to the type of answer expected. This figures fall to less than 50% for the medical type. Evaluations have showed that the system was good to identify the type of answer expected and could be enhanced for the medical type. It leads us to use an external source of knowledge: UMLS. A future improvement will be the usage of UMLS semantic network to better categorize a query according to the medical domain.

Keywords. Model detection, Supervised classification, Medical Type

Introduction

The development of Internet places, an increasing quantity of information at our disposal. The main problem resides in the access to this information. How much time is needed for you to find the precise answer to your question? Indeed, search engines optimised for query keywords, only provide a long list of Web documents, lists that one necessarily has to browse to find an accurate answer. The alternative to this difficulty is a question/answering system. The goal of such system is to extract precise answers to a question submitted in natural form rather than a list of documents. Its role is clearly to help users as quickly as possible towards access to required information instead of having to look through all documents proposed by search engines.

With the progress in the medical field and better access to healthcare, the average life span is constantly growing. Along with this grows the interest and concern the average person shows towards information regarding his health, thus creating a bigger

demand for information on the Internet. The medical domain is widely represented on the Web but how can a common user judge the relevance of medical documents?

It is to answer this double problem that the Health On the Net Foundation (HON), a leading organization in promoting and guiding the deployment of useful and reliable online medical and health information became involved in the development of a question/answering system specific to the medical domain, which is a complex domain requiring the limitation of the framework of research. This research activity is done within the European project PIPS (Personalized Information Platform for life and health Services). The PIPS Project has the main aim to create a new Health and Life Knowledge and Services Support Environment for protecting the health of the Individual. PIPS is an advanced platform for health personalized management using only quality knowledge sources. QA system is an alternative to usual search engine and can fill the gap of the PIPS ontology server and semantic engine which is limited to few domains and not exhaustive.

There are no QA systems applied to health available on the Internet but only studies found in scientific literature [1, 2, 3, 4, 5]. However, by the 1960s, people were already attempting to come up with a direct answering system to a question submitted in a natural form. Two of the most famous QA systems are BASEBALL (questions on the US baseball league) and LUNAR (on geological analysis of rocks). Nowadays, general QA systems are available on the Web like AnswerBus or BrainBoost [6].

Most QA systems use pattern matching to classify sets of questions [7, 8]. However machines learning is known to be capable of identifying data an expert would not have thought of, whatever the language [3, 9]. The questions classification is the main issue in QA system using machine learning. Annually competitions like TREC or CLEF aim to evaluate such systems.

The application developed by HON is available in both English and French. In addition, the user is able to choose the domain of research i.e. Websites accredited by HON or all the websites, regardless of accreditation. By default the research is based on the accredited sites as the quality of the available information is paramount, far

more so than the quantity. Figure 1 presents the architecture of the QA system. It is composed of five modules. The module “QuestionAnalyser” takes place at the beginning of the QA system. Figure 1 shows the importance of this module for the continuation of the treatment of a question. Indeed, all modules describe in figure 1 use its results to carry out their treatment to seek and retrieve the answer.

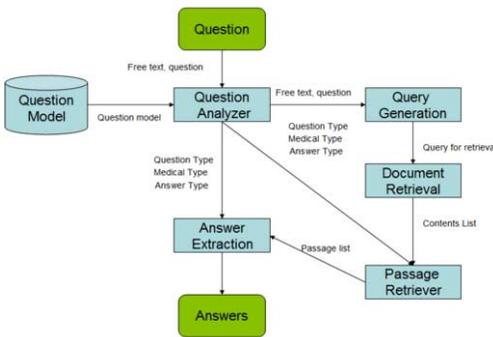


Figure 1. Q&A system architecture

Our hypothesis is that its medical type and its type of expected answer characterize a medical question.

The following sections present the material and methods used to classify medical questions and the results obtained for the supervised method.

1. Materials and methods

1.1. Corpora

Two corpora were used for this study: the first was composed of 136 questions in English and the second, 140 questions in French. In both cases, the questions were collected from Frequently Asked Questions (FAQ) found in forums specialized in health as well as discussions on the Internet. We tried to retrieve questions covering as much as possible of the wide spectra of the subject within the topic. The medical field being very vast, it was agreed to focus only on the topic diseases. The project PIPS targeting the diabetic's patients, and the module of analysis of questions carried out was also targeted on this disease. We make the assumption that this module is independent of the type of medical subject. The choice to study only one topic makes it possible to only categorize the relevant classes and not the medical topic.

1.2. Non supervised method

In a first step we preferred to classify automatically questions, with no information about their categories. For this task we used the software Rapid Miner which is an environment for machine learning. It consisted of automatically detecting clusters only from the corpora according to the characteristics of the questions. Moreover, according to the algorithm used, it is possible to fix or remove the number of classes of exits.

Initially, the questions were presented in the untreated form (called unigram form). Several algorithms of clusterisation were tested (K-Means, EM, DBSCAN, Kménoids, and Kernel-KMeans).

Following this, questions were proposed with bigrams of words or co-occurrences. The bigrams of words allow identifying the expressions composed. The co-occurrences associate the distant words within a sentence belonging to the same linguistic unit. It is considered that there is co-occurrence when the presence of a word in a text gives an indication to the presence of another word. The interest to apply such work to the set of questions is to locate the expressions, associations of words, which are recurring on the entire corpus.

Non supervised classification did not emphasize new classes of questions. However, we could observe that the software creates clusters according to the type of interrogation term. For instance, all questions beginning by "What is" were in the same cluster. Consequently, this work was not useless and has prepared the step of conception by providing distinctive structures of sentences and useful expressions.

1.3. Classification by an expert

The two sets of questions used for this study were provided by two physicians who classified them in two different ways: according to the type of answer expected and

Table 1 and Table 2. Question repartition for the medical types and the types of answer expected.

	En	Fr
Causes	8	9
Diagnostic	6	2
Diet	18	14
Disease	25	12
Evolution	12	11
Physiology	15	11
Prevention	6	8
Routine	10	34
Symptoms	13	8
Treatment	32	55

	En	Fr
Boolean	53	65
Causal	17	5
Definition	47	12
Factoid	6	13
Last	4	3
List	7	15
Moment	4	6
Person	7	4
Place	9	4
Procedure	4	30
Quantity	15	7

according to the medical type of the question. For the medical questions, the medical expert classified the questions according to the medical language used, while focusing on the medical meaning of the words. The classification by categories suggested corresponds to topics proposed in medical text books.

Table 1 and Table 2 give this manual classification. It has brought out 10 medical types and 11 answer types, categories for both English and French.

According to human experts, whatever the language considered, questions most usually asked are related to the treatment of a disease. For the type of expected answer, Table 1 and Table 2 show that the most representative categories are “Boolean” and “Definition”. It is due to the data-gathering which is based on questions most frequently asked with no consideration of proportion (actually categories were unknown during the data-gathering).

The characterization according to the type of answer expected found in our research matches the one described in the literature [10, 11, 12]. However we decide to re-define categories because of the specificity of our study which is based on medical questions only.

1.4. Supervised method and criteria of judgment

Questions were gathered according to the expert classification. The tool learns according to this classification and generates a model. This model was applied to the same set of questions, following which cross validations were carried out. The five algorithms tested are those found in the literature for the task of classification [3]: SVM, NaiveBayes, Knn, ROCCHIO, Decision Trees.

This method was carried out in two phases: according to the medical categories and the type of expected answers. In addition, criteria of judgment were used to apply cross validation. The corpora were divided into two sets: the training base (90% of the corpus) and the one of tests (10%). Cross evaluation have been realised. This was done to allow us to calculate a score of performance of the automatic tools and its relevance for the task. Seven measures were used: macro (ma) and micro (mi) recall(R: Capacity of the system to report only relevant documents)/precision(P: Rate of relevant documents proposed by the system as compared to all the retrieved documents)/F1 and

Table 3. Summarization of the classifiers' results

	<i>maR</i>	<i>maP</i>	<i>maF1</i>	<i>miR</i>	<i>miP</i>	<i>miF1</i>	<i>Erreur</i>
English Med type (SVM W2*)	0.47717	0.52499	0.476617	0.69446	0.410362	0.410362	0.078078
English Ans type (SVM W2*)	0.80527	0.84569	0.81519	0.77891	0.91934	0.84031	0.026561
French Med type (SVM W2*)	0.47505	0.51217	0.46431	0.312687	0.62901	0.41354	0.086006
French Ans type (NB W3*)	0.7744	0.68595	0.68028	0.71381	0.63112	0.66846	0.066063

* Wx = x grams of words, Med = Medical and Ans = Answer

the error. Macro-measures compute values of precision/recall for each category and make a mean on these values. Micro-measures gather data of each category in a same contingency table and compute values of precision/recall according to this table. The F1-measure is a function that is maximized when the precision and the recall are near. We attached more importance to the values of precision than to those of recall because it is necessary that the composition of the categories proposed by the classifier coincide with that of the human experts. Finally, this set of measures has been calculated according to units of treatment (unigram, bigrams, trigrams and co-occurrences of words).

2. Results and discussion of the supervised classification

The tests of supervised classifications have made it possible, both to validate the categories of questions provided by the expert and to determine the best algorithm for the step of implementation. It has been realized according to both categories for medical types and categories for the type of answer expected.

The results show that a question is better classified if it belongs to a category well represented in the initial corpus. For instance, 91 % of "Definition" questions (34.3% of the English's corpus) are well-classified with the NaiveBayes classifier whereas it falls to 60% for the category "Procedure" (11.2% of the English's corpus).

Table 3 gives results for the medical type and the type of answer expected, with the best classifiers and units of treatment. As results found in literature [13], the two algorithms which are most relevant for the task of classification are NaiveBayes and Support Vector Machine (SVM) with almost 85% of micro-precision for the type of answer expected in English. Moreover, results are better for the type of answer expected. Indeed the values of recall and precision are around 80% against less than 50% for the determination of the medical type of the question (cf table 3). It can be explained by the lack of data for the medical types. Furthermore, the number of questions by category is better distributed for the type of answer expected. Actually some categories are very poor. For instance, "Diagnostic" is only represented by 2 questions in the French's corpus. In addition, this imbalance can be explained by the complexity to establish a model for the medical types. Two questions having the same category for the type of answer expected have a common sentence structure. It is not the case for questions belonging to the same medical category. For instance the two questions following belong to the category "Procedure" for the type of answer expected: "**How can** diabetes affect my mood?" and "**How can** I take care of myself if I have diabetes?". They have a common structure: "How can + subject + verb+

complement”. Let’s take two questions belonging to the medical category “Cause”: “Am I at risk for diabetes?” and “How do I know if my kidneys are **affected** “. There are no commons structures. So the machine learning task is more difficult for categorizing the medical type.

Globally, the automatic classification is better for English than for French. Table 3 shows that for the type of answer expected the system is capable of 84% for English and 68% for French (macro-precision). It can be explained by the linguistic differences between French and English (e.g. conjugation).

3. Conclusion and Perspectives

For the realization of a QA system applied to health we decided to use a supervised method to characterize questions whereas most of the QA systems are using pattern matching. From set of questions and with the expert classification we have proposed a new taxonomy of clinical questions for QA systems whereas the non supervised classification of this set of questions was not efficient.

The evaluation of the Question Analyzer module for the task of classification confirms the feasibility of the supervised method for a QA system for both English and French. Indeed, SVM classifiers are highly capable of performing the task of classification according to the type of answer expected.

The careful analysis of results obtained and errors explains that the less well recognised classes are the one that are less well represented in our learning corpora. Thus we have to enhance our training base for both English and French. And future research which overcome results obtained by the determination of the medical type will be focussed on the usage/integration of the Unified Medical Language System (UMLS) network. This language provides precise relations between medical terms and many synonyms for each of them.

References

- [1] P. Zweigenbaum, Question answering in biomedicine. EACL workshop on natural language processing for question answering, Budapest, 2003, p. 1-4.
- [2] M. Lee, J. Cimino, H.R. Zhu, C. Sable, Y. Shanker, J. Ely, et al. Beyond information retrieval - medical question answering. AMIA Annu Symp Proc (2006), 469-473.
- [3] H. Yu, C. Sable, H.R. Zhu, Classifying medical questions based on an evidence taxonomy, In: Proceedings of the 2005 AAAI Workshop on Question Answering in Restricted Domains, Pittsburgh, PA, 2005.
- [4] T. Delbecque, P. Jacquemart, P. Zweigenbaum, Indexing UMLS semantic types for medical question-answering, Stud Health Technol Inform **116** (2005), 805-810.
- [5] Y. Niu, Question answering in medicine, Technical Report for the Oral Qualification Exam, Department of Computer Science, University of Toronto, 2003.
- [6] AnswerBus: <http://www.answerbus.com/index.shtml>; BrainBoost: <http://www.brainboost.com/>.
- [7] P. Jacquemart, P. Zweigenbaum, Towards a medical question-answering system: a feasibility study, Stud Health Technol Inform **95** (2003), 463-468.
- [8] E. Alfonseca et al, A prototype question answering system using syntactic and semantic information for answer retrieval, Proceeding Text Retrieval Conference, 2002.
- [9] R. May, A. Steinberg, Building a Question Classifier for a TREC-Style Question Answering System, The Stanford Natural Language Processing Group, Final Projects 2004.
- [10] F. Benamara, P. Saint-Dizier. Construction de réponses coopératives. Du corpus à la modélisation informatique. Revue Québécoise de Linguistique **18** (2004) (3), 34-59.
- [11] S. Mendes et al. L’analyse des questions : intérêts pour la génération des réponses, TALN, in: TALN- Workshop question-réponses, Fes (Maroc), 22/04/04-22/04/04, Association pour le Traitement Automatique des Langues (ATALA), 2004, p. 11-22.
- [12] O. Ferret et al. How NLP can improve question answering. Knowledge Organization **29** (2002) (3-4), 135-155.

Group Decision Support System applied to the medical pluri-disciplinary decision group: Usability and Efficacy

Nathalie DEGARDIN-CAPON, Nathalie BRICON-SOUF, Marie-Catherine.

BEUSCART-ZEPHIR, Régis BEUSCART

CERIM-EVALAB –EA2694- Faculté de Médecine – Université de Lille2 –Lille (France)

Abstract . Objectives: This paper aims to study whether the application of a Group Decision Support System to medical collective decision committees is possible and to determine which GDSS specifications are convenient.

Backgrounds: we introduce the common knowledge about GDSS and define the process of the collective medical decision. **Methods:** An experimental GDSS has been tested in an actual medical collective decision committee. A usability analysis has been performed to precise usability and acceptability of the system and to highlight pro and cons of the various functionalities of the GDSS. **Results:** Information sharing was conveniently supported by the GDSS. All the documents were available for the support of the discussion. But, the introduction of a GDSS in the decision committee added new constraints such as the necessity of an excellent preparation phase. Limits of the system have been revealed: lack of feedback on decision actors, lack of support to obtain the consensus and lack of memorisation. According to these results, we have proposed new GDSS features to improve the decision. **Discussion-Conclusion:** Using a GDSS supporting the medical collective decision is realistic and may support the process of the consensual decision.

KEY-WORDS: group decision, medical decision, GDSS, collaborative work, decision committee

Introduction

Medical collective committees are necessary to manage patients with complex or specific pathologies, needing experts' consensus [1]. They are mandatory in specific medical domains as oncology where inter-disciplinary cooperation is essential in the management of the decision process.

Last twenty years, the development of computer technology permitted the emergence of the concept of Group Decision Support System (GDSS). The features proposed by the GDSS (information sharing, decision process support, communication support, group memory) could emphasize group decision to achieve consensual decision. Thus, applying GDSS features to the medical pluri-disciplinary decision committees appears to be an interesting investigation way [2].

The objective of this work is to experiment and evaluate the usability and efficacy of a GDSS prototype in an urological oncology decision committee.

1. Background

GDSS: The concept of GDSS is issued from the confluence of two fields of the Computer Science: the Computer Supported Co-operative Work (CSCW) systems and the Decision Support Systems (DSS) [3]. GDSS is defined as “an interactive computer-based system that facilitates the solution of unstructured problems by a group of decision makers” [4]. Nowadays, the use of a GDSS is widespread in business, government, military and professional services sectors [5]. The traditional concept of GDSS is the decision room or face-to-face GDSS, for real time decision making in a same time meeting [4]. These are electronic meeting rooms containing a large U shaped table with embedded computers linked through a local area network (LAN), a projection system and a large public screen. Multiple projection devices allow interface with simultaneous information sources such as a video teleconference or web pages. The meeting is usually initiated by a human facilitator whose function is to guide the group and to act as an interface between the group and the devices. Participants, independent of each others provide simultaneously and anonymously the system with ideas or propositions. GDSS typically supports some aspects of the decision process such as brainstorming, idea organization, evaluation, prioritization, and voting. Vote is the mean used to reach the consensual decision. At the end of the meeting, the results of the information captured by the system can be delivered in a formatted report to the participants for immediate action.

GDSS attempts to address several negative aspects of the face-to-face group process, such as dominance of discussion by one participant, influence of high status members of the group, low tolerance of minority opinions, pressures toward group consensus and problem of coordination.

Collective medical decision

The collective medical decision is a very complex task and a pluri-disciplinary process [1]. The medical collective decision committees are mandatory in case of complex pathologies, needing different expertises or therapeutic processes [1]. For fifteen years, these decision committees have been more and more developed to assist the management of rare diseases (often concerning people with congenital malformations or dysfunctions) or life-threatening pathologies (cancer), requiring the cooperation of different experts issued from various medical fields.

During medical collective decision making, experts from different medical fields, mostly in a “same place-same time meeting”, share information and discuss to find a consensus: a common diagnosis and a common therapeutic decision. Medical information is shared through text and images. Experts comment and discuss the documents to refine the diagnosis. They have to discuss and to agree about a common decision of treatment. The discussion is often promoted by a leader (an elder or more graduated and experienced expert).

2. Materials and Methods

2.1. Materials

We have studied a medical collective decision committee devoted to the management of patients suffering from urological tumors. This urological oncology consultation (UOC) was assisted by an experimental GDSS (UOC-GDSS).

Description of the decision committee

The urological oncology decision committee (UOC) gathers different experts: urology surgeon, histologist, radiologist, oncologist and students (residents). Meeting takes place once per month. All the discussions are based on various medical and paramedical data issued from numerous pre-existing examinations. Information is made of multimedia data: histological images, CT scans and X-Rays, specific textual summaries of the clinical data.

Scenario of the medical collective decision .

During a session, the experts are around a table; they have access to common documents. Each expert orally presents or comments the exams of his/ her speciality; The experts share the various medical documents, discuss then, and establish a consensus to propose the best decision for the patient.

During this procedure, a physician plays the role of a scrivener. His task is to write on a paper the elements of decision and the final decision issued from the collective discussion in order to elaborate the meeting summary.

2.2. Description of the prototype of GDSS

This system has been implemented on the concept of a local server based upon the peer-to-peer principle: the system checks patient's documents on different servers of the hospital Information System (X-rays, histological report, letters, lab results...). If a document for a patient is found, it is then downloaded in the GDSS system for possible display. (**Figure 1**)

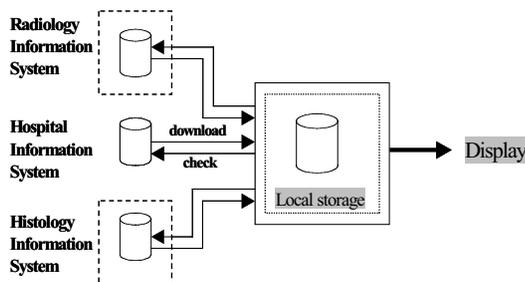


Figure 1: The architecture of the GDSS Prototype

During the meeting, the pilot chooses to display them for presentation, for expert's comments and for discussion. Three different documents can be simultaneously displayed on a shared public screen.

2.3. Methods

Two usability engineers trained in cognitive psychology and ergonomics from the Usability Lab “EVALAB” [6] have realized the “task and activity analysis” in the OUC. The methodology of the study was based on (i) repetitive natural observations, (ii) audio and video recorded observations, (iii) analysis of tracks of the activity and (iv) user interviews. All participants in the project were given a full report of the task and activity analysis’ results.

Natural Observation of the meeting

4 meetings were observed. A meeting gathered at least five participants: one surgeon, one oncologist, one radiologist, one histologist and one student.

Interview of the participants

8 usual participants of the meeting have been observed and interviewed: 3 urologists, 1 oncologist and 1 radiotherapist, 2 radiologists, 1 histologist, 1 resident. The medical secretaries of each involved department have also been observed and interviewed: they do not intervene during the meeting process but during the preparation task of the meeting. User interviews were individually conducted.

Activity evaluation in the Usability Lab

An activity evaluation of the different members of a multi-disciplinary meeting assisted by the UOC- GDSS was conducted integrating usability engineering methods [6] in the usability lab. 4 participants were involved in this test: one surgeon, one radiologist, one histologist and one resident.

3. Results

3.1. Usability study

From the usability study, we obtained the following results. Two successive phases compose the decision process: the preparation task and the decision meeting itself see (Figure 2).

The Preparation task

This first consists, in opening a new meeting session: (date, list of patient’s name). The collection of medical is mainly realized by the medical secretary. Unlike the common idea (4/8 of the interviewed physicians) that the preparation task is time- consuming, the secretaries found this phase really efficient as this task did not exceed 15 minutes.

The meeting takes place one week later.

It is a same time, same place meeting with a pluri-disciplinary expertise (small group of 4 to 9 participants). From the observations realised in the Usability Lab, it was demonstrated that the GDSS prototype fulfilled the main functions required by the users for the realisation of a completed session of pluri-disciplinary decision meeting.

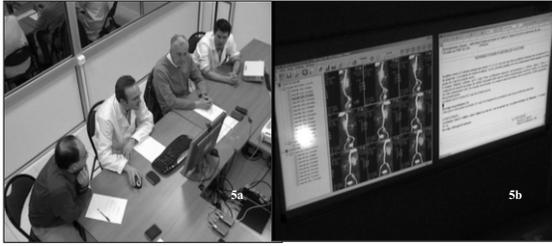


Figure 2: Experts using the GDSS in an experimental meeting in the Usability Lab.
 On the left: a resident plays the role of pilot. The 3 experts focus on discussion and decision tasks.
 On the right: visualisation of the documents on a public screen.

Display of documents: The modalities for displaying documents have to be discussed: which documents to display? In which format? It was proposed to display document in the 100% format and to introduce a special procedure to give a rapid access to various documents when they are too many.

Errors: the patient’s identity must be closely linked to the different documents to avoid confusion amongst the documents of various patients.

Flexibility: the display of documents must be as flexible as possible so that the experts can have a simultaneous access to different elements of the record and display them on different parts of the screen.

“Private” space for the radiologist: the radiologist wants to have a “private session” to prepare the images to be displayed for the other participants. This could avoid lack of time and d better choice of the relevant images.

Generalization: Currently, the system is available in special rooms of the University Hospital. To be successful, the application should be available in all the decision rooms within the hospitals.

3.2. Benefits and Limits

More generally, the main benefits, limits and constraints of the system are summarized in table 1.

Benefits
Equal access to the information for each participant by a public shared screen.
All the documents are available for the support of the discussion
Experts are free of any manipulation thanks to the “pilot” <i>(thus they can remain focused on the discussion and decision task)</i>
Limits and constraints
Necessity of an excellent preparation
No memorization of the results of the decision
Lack of feed back and gap of time for integrating the new information

Table 1: Benefits, limits and constraints of the introduction of the UOC GDSS in the urological oncology decision committee

4. Discussion and Conclusion

GDSS are gaining widespread acceptance. Business managers spend as much as 70 to 80 percent of their time in meetings. GDSS improve the process of the meeting by quickly generating, organizing and evaluating large amounts of information.

Medical collective decision committees are more and more often required in the medical domain to help the management of patients with complex or specific pathologies needing experts' consensus.

We wondered whether the application of a GDSS concept to a medical pluridisciplinary decision committee, could be acceptable and useful for the medical staff.

The proposed experimental GDSS has been well accepted and has successfully supported the information sharing. In fact, GDSS propose many features that seem to be convenient to the different phases of the medical collective decision process.

But the current system is unperfect and our study has pointed out some limits: this has led us to propose a refined model of medical collective decision supported by a GDSS and to define some new specifications for the medical GDSS. Our development perspectives are now to improve the GDSS prototype, while being restricted with the appropriate features for the medical collective decision.

Nevertheless, new prospects remain not considered by our study of the existing. One could for example, take into account some possibilities of innovation, such as access to the Web, use of mobile tools, integration in the HIS...etc...The GDSS could allow to enlarge the possibilities of organisation of medical collective decision meetings. For example, supporting the meeting by a Web-based GDSS could make it possible for other distant experts to join the meeting without any displacement or to imagine an asynchronous mode of decision in order to avoid the problems of agenda synchronization.

Thus, a medical GDSS consisting in (Web-published) shared medical data bases, public visualization features, easy access to documentation tools, mobile voting system (avoiding the unlike keyboard's use) to support consensual decision and memorisation features appears to be an interesting and realistic development project to support the medical collective decision.

References

- [1] Sauvagnac C, Beckendorf V, Lesur A, Luporsi E, Stines J, Falzon P, Bey P Decision committees in oncology: from standard to actual cases. *Bull Cancer*. Number 86, volume 9, 767-72, September 1999
- [2] GR Rao, M Turoff. A hypermedia-based group decision support system to support collaborative medical decision making. *Decision Support Systems* 30(2000)187-216
- [3] Ellis CA, Gibbs SJ, Rein GL. Groupware, Some issues and experiences. *Communications of the ACM*, vol.34, n°1, 38-58, January 1991.
- [4] Ellis CA, Gibbs SJ, Rein GL. Groupware: some issues and experiences. *Communications of the ACM*, vol.34, n°1, 38-58, January 1991.
- [5] ShenQ, Chung J, A group decision support system for value management studies in the construction industry. *International Journal of Project management* 20(2002)247-252
- [6] Beuscart-Zephir, M. C., Anceaux, F., Menu, H., Guerlinger, S., Watbled, L., & Evrard, F. (2005). User-centered, multidimensional assessment method of clinical information systems: a case study in anaesthesiology. *International Journal of Medical Informatics*, 74, 179-189.

The Gap between Actual and Mandated Use of an Electronic Medication Record Three Years after Deployment

Maren Fich GRANLIEN^{a,1}, Morten HERTZUM^a, Jette GUDMUNDSEN^b

^a *Computer Science, Roskilde University, Denmark*

^b *Region Zealand, Denmark*

Abstract. Three years after the hospitals in one of Denmark's five healthcare regions deployed an electronic medication record (EMR) four of eight main system facilities are used consistently by only 3%-37% of the hospital wards. Furthermore, four of eight mandated work procedures involving the EMR are followed consistently by only 13%-28% of wards. No system facility or work procedure is consistently adopted by more than 67%, respectively 48%, of wards. Barriers to adoption of the EMR include system factors, such as the EMR being perceived as prohibitively time consuming to use, as well as human factors, such as lack of knowledge, information, and training among clinicians. However, the prime barrier appears to be uncertainty about what the barriers concretely are and about the extent to which system facilities and work procedures are actually adopted. Three years after deployment it is apparent that time alone does not lead to consistent adoption. Rather, interventions are necessary to overcome the barriers.

Keywords. EMR, implementation, deployment, diffusion, compliance.

Introduction

As part of the extensive efforts to substitute electronic patient records for paper records at Danish hospitals, Region Zealand (one of five healthcare regions in Denmark) started deploying an electronic medication record (EMR) at the hospitals in 2003 and finished deployment in early 2006. The EMR is now used on all hospital wards and in some out-patient clinics by approximately 10000 physicians and nurses for maintaining an overview of patients' medication. The physicians also use the EMR for ordering medication and the nurses for dispensing and administrating medication. The intention of the EMR is to help ensure that the right medication is given to the right patients at the right time. Several work procedures involving the EMR are mandated in the region's standard operating procedures for medication. The EMR is part of the electronic patient record, and the implementation of the EMR and the associated work processes and clinical guidelines aims to improve patient safety and documentation quality. It is, however, a general impression in Region Zealand that these aims have not

¹ Corresponding author: Maren Fich Granlien, Roskilde University, bldg. 42.2, Universitetsvej 1, DK-4000 Roskilde, Denmark. E-mail: maren@ruc.dk

been attained and that the EMR is not used as intended. Such gaps between an organizational decision to acquire a system and the actual use of the system by people in the organization have been termed assimilation gaps [1]. Assimilation gaps indicate that the actual use of a system is a separate decision, not simply a product of the decision to acquire the system. In this study we analyse the extent to which the different parts of the EMR are used and the extent to which the mandated procedures for using the EMR in the medication process are followed. Our interest in how profoundly the EMR has become incorporated in the practices of the hospital staff is motivated by a belief that “for a technological innovation to be truly valuable, it must be incorporated within the adopting organization’s operational or managerial work system” [2].

The aim of this study is twofold: (a) to assess the actual adoption of the EMR among hospital staff and their compliance with guidelines for its use and (b) to investigate possible barriers toward adopting the EMR and associated work procedures after the hospital staff had gained considerable experience with the system.

1. Method

The data for the study were collected by means of an online questionnaire that was developed by the authors and administered with the survey tool SurveyXact®. An email requesting participation was sent to all function managers, department managers, ward managers, and EMR coordinators at the hospitals in Region Zealand, a total of 430 people. Participation in the survey was anonymous and after issuing two reminders we received 232 responses (94 physicians, 129 nurses, 9 others), for a response rate of 54%. The targeted clinicians were managers at the mid and lower levels. We targeted these clinicians for two reasons. First, contrary to end-users the targeted clinicians could answer on behalf of the entire unit for which they were responsible. Second, if the targeted clinicians’ answers revealed uncertainty about the actual adoption of the EMR this was itself interesting because these clinicians should according to organizational guidelines be able to answer the questions on behalf of their unit.

Questions in the survey concerned the adoption of the EMR and associated work procedures. Respondents were asked to what extent different parts of the EMR were used and to what extent different work procedures were followed. The response categories for these questions were *Always*, *Very often*, *Often*, *Rarely*, *Very Rarely*, *Never*, and *Don’t know*. Respondents were also asked to describe, in free text, perceived barriers to using the facilities of the EMR and complying with the work procedures. Moreover, information about the training provided in the use of the system was elicited through a number of questions not analysed in this paper. The questionnaire comprised 59 questions in total and was estimated to take 15 minutes to complete.

Respondents provided 522 free-text comments about barriers to the adoption of the EMR and associated work procedures. These comments were analysed and categorized by the first and third authors through a collaborative process of affinity diagramming [3]. To assess the reliability of the resulting 12 categories, the second author independently assigned each comment to a category. The Kappa value for the level of agreement between the two categorizations of the comments was 0.72, which according to Landis and Koch [4] corresponds to ‘substantial’ agreement. Disagreements were resolved through discussion and a consensus was reached.

2. Results

Table 1 shows the extent to which the main facilities of the EMR are used by the wards of the hospitals in Region Zealand. Though the system was designed to support the clinicians' work, none of the facilities are used always or very often by more than two thirds of the wards, and four of the facilities are used always or very often by only 3%-37% of wards. This partial adoption of the system facilities is particularly noteworthy for the three facilities, the use of which is mandated in the region's standard operating procedures for medication. Furthermore, the extent to which one system facility is used at a ward weakly indicates that the other system facilities are used to a similar extent in that the average pair-wise Spearman correlation between system facilities is 0.30 ($SD = 0.15$), $p < 0.05$ for 24 of the 28 pairs of correlation.

Questions about the use of five of the system facilities (questions 1, 3, 6, 7, and 8 in Table 1) were also included in a survey in 2004 at the first hospital that deployed the system. A comparison of the responses indicates that the extent to which these five system facilities are used always or very often has increased about 20 percentage points in the three years since 2004.

Table 1. Extent to which system facilities are used, $N = 232$.

Use of system facilities (tab sheets)	Always + Very often	Often + Rarely	Very rarely + Never	Don't know	Mandated
1. Overview (of ordered medication and its dispensing/administration)	56%	19%	3%	22%	
2. Medication orders	64%	15%	7%	14%	Yes
3. Regimens	37%	39%	13%	11%	
4. Medication history	9%	49%	18%	24%	
5. Prescription history	3%	28%	33%	36%	
6. Prescriptions	21%	37%	25%	17%	
7. Dispensing/administration ^a	67%	4%	11%	19%	Yes
8. Dispensing/administration ^b	63%	8%	10%	18%	Yes

^a When medicine is dispensed, ^b When medicine is administered.

Table 2 shows the extent to which work procedures involving the EMR are followed. Apart from the use of standard medication orders all these work procedures are mandated in the region's standard operating procedures for medication. However, none of the work procedures are followed always or very often by more than 48% of wards, and four of the nine work procedures are followed always or very often by at most 28% of wards. The extent to which one work procedure is followed at a ward weakly indicates that the other work procedures are followed to a similar extent in that the average pair-wise Spearman correlation between work procedures is 0.29 ($SD = 0.22$), $p < 0.05$ for 25 of the 36 pairs of correlation.

Respondents were also asked to indicate the overall extent to which the standard operating procedures for medication were complied with. Though answers to this question correlated significantly with answers to six of the nine questions about the

extent to which specified work procedures were followed the correlations were weak, suggesting limited awareness of the content of the standard operating procedures. Furthermore, many respondents lacked knowledge of the extent to which specified work procedures were followed, as indicated by the high percentages of *Don't know* answers. Averaged over all nine work procedures, 31% of function managers ($N = 35$) and 30% of department managers ($N = 79$), the subgroups of respondents formally responsible for their unit's compliance with standard operating procedures, gave *Don't know* answers. Across all 232 respondents the average was 26%. The percentage of respondents uncertain about the extent to which system facilities were used was slightly lower, but still averaged 20% across the eight questions in Table 1.

Table 2. Extent to which work procedures are followed, $N = 232$.

Work procedures	Always + Very often	Often + Rarely	Very rarely + Never	Don't know	Mandated Compliance assessment ^a
1. Use of standard medication orders	43%	38%	8%	11%	0.07
2. Dispensing of each medicine is signed for separately	48%	14%	8%	30%	Yes 0.14 *
3. Administration of medicine is signed for when it is given to patient	34%	25%	10%	31%	Yes 0.15 *
4. Medicine status is set when a patient is admitted	34%	39%	9%	19%	Yes 0.32 ***
5. Medicine status is set when a patient is transferred	28%	42%	9%	21%	Yes 0.27 ***
6. Medicine status is set when a patient is discharged	38%	34%	8%	20%	Yes 0.38 ***
7. Administration status is set when a patient is admitted	27%	36%	8%	29%	Yes 0.22 **
8. Administration status is set when a patient is transferred	13%	31%	15%	41%	Yes 0.10
9. Administration status is set when a patient is discharged	19%	34%	12%	35%	Yes 0.09

^a Spearman correlation with question 'Standard operating procedures for medication are followed', * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3 shows the five categories of barrier most frequently mentioned by respondents in their free-text comments, totaling 73% of the comments. The categories can be divided into those related to system factors and those related to human factors. System factors include poor usability and overview (e.g., "Difficult to get an overview due to an illogical composition of the interface") and areas inadequately supported by the system (e.g., it is difficult to handle infusion medicine in the EMR because the frequent adjustments of the infusion rate are cumbersome). Time is a system factor in the cases where it refers to the slow response times of the EMR or when inferior design is the reason for the system being time consuming to use. While most of the comments

about time refer to system factors, time can also be a human factor. This is, for example, the case when insufficient computer skills and lack of training are the reasons why system use takes a lot of time. The categories related to human factors include lack of knowledge, information, and training, but the largest human-factors category is uncertainty about what constitutes the barriers to using different parts of the EMR.

Table 3. The five most frequent categories of barrier, $N = 522$.

	Category	Number of comments
1	Don't know: stating that barriers exist but not knowing what they are	132
2	Time: the system being too slow and time consuming to use	85
3	Lack of knowledge, information, and training	60
4	Inadequate support of certain work areas	55
5	Poor usability and overview	50

3. Discussion

Three years after deployment there is a considerable gap between mandated and actual adoption of the EMR and associated work procedures. This gap exists in spite of several attempts during the past three years to address some of the barriers toward using the system and complying with standard operation procedures for medication. Both Region Zealand and the EMR vendor have been aware of the slowness of the system and have tried to improve the network, the computers, and the design of the EMR itself. Furthermore, the regional implementation organization has established a standard training program for new staff, and continuously throughout the last three years extra information and training have been provided.

Function and department managers (49% of the survey respondents) are formally responsible for their unit's consistent use of the EMR and compliance with medication procedures. Hence, these respondents ought to know the extent to which the EMR and work procedures are adopted, the barriers that impede consistent adoption, and how to address these barriers. The responsible managers are, however, to a considerable extent uncertain about what the barriers concretely are, complicating directed efforts to address the barriers. The remaining respondents display a similar uncertainty.

Several concrete barriers are, however, mentioned frequently by respondents. Though it may in some situations be difficult to determine whether time barriers are due to insufficient hardware, inefficient EMR software, or inadequately trained staff it is unquestionable that time is perceived and pointed out as a vast barrier by the respondents. It therefore has to be addressed. Similarly, the lack of knowledge, information, and training needs to be addressed. The previous attempts at increasing knowledge and information through training serve to emphasize that this barrier has to be addressed in a more targeted, effective, and systematic manner to achieve adoption. A recent study of another EMR implementation in Denmark showed that three months after deployment the actual level of use was far below the desired level of use. This assimilation gap was partly explained by the short period of use [5]. In our study we

find a similar gap after three years of use, suggesting that it may be overly optimistic to expect that a long period of use will lead to a gradual closing of such gaps. A candidate explanation for the persistence of assimilation gaps is the concept of media stickiness [6]. Huysman et al. [6] found that the patterns of use for a new system are developed shortly after the system is deployed and tend to persist over time.

We see a need for an increased focus, by the responsible clinical managers as well as by the EMR implementation organization, on the barriers toward using the EMR and on ways of addressing these barriers. To overcome the barriers, this focus must be accompanied by activities to monitor whether the interventions that are undertaken have any effects and, if not, by the launching of new interventions. It would also be desirable with a more in depth ethnographic study to unveil unidentified barriers.

4. Conclusion

In this study we have found a considerable gap between mandated and actual use of an EMR and a lack of compliance with the work procedures associated with the EMR. These findings are not a result of limited experience with the EMR but the state of affairs after three years of use. The EMR is fully diffused at the organizational level but at the level of clinicians the adoption of the EMR and its incorporation into clinicians' work practices are far from the level necessary to attain the goals that motivated the acquisition of the EMR. A number of barriers related to both human and system factors enter into explaining the gap, but considerable uncertainty also exists about what the concrete barriers actually are. We therefore suggest further investigation of the barriers and experiments exploring what kinds of intervention can counteract the barriers. This study indicates that time alone will not lead to consistent adoption; hence, we cannot expect the EMR to be used as mandated unless the barriers are addressed. Hence, better knowledge of existing barriers and effective interventions are required.

Acknowledgements

We thank Mikkel Lundstrøm, who was instrumental in the administration of the survey. Special thanks are due to the survey respondents.

References

- [1] Fichman, R., G. and C. Kemerer, F. , *The Illusory Diffusion of Innovation: An Examination of Assimilation Gaps*. Information Systems Research, 1999. **10**(3): p. 255-275.
- [2] Zmud, R.W. and L.E. Apple, *Measuring Technology Incorporation/Infusion*. Journal of Product Innovation Management, 1992. **9**(2): p. 148-155.
- [3] Beyer, H. and K. Holtzblatt, *Contextual design : Defining Customer-Centered Systems*. Morgan Kaufmann series in Interactive Technologies, ed. S. Card, Grudin, J., Linton, M., Nielsen, J., Skelly, T. 1998, San Fransisco: Morgan Kaufmann.
- [4] Landis, J.R. and G.G. Koch, *The measurement of observer agreement for categorical data*. Biometrics, 1977. **33**(1): p. 159-174.
- [5] La Cour, V. and C. Hellstern-Hauerslev, *Evaluating the implementation and use of a computerized physician order entry system: a case study*. Stud Health Technol Inform, 2007. **130**: p. 75-9.
- [6] Huysman, M., et al., *Virtual Teams and the Appropriation of Communication Technology: Exploring the Concept of Media Stickiness*. Computer Supported Cooperative Work: The Journal of Collaborative Computing, 2003. **12**(4): p. 411-436

Analysis of EHRs for research, quality management and health politics

Walter GALL,^{a,1} Wilfried GROSSMANN,^b Georg DUFTSCHMID,^a
Thomas WRBA,^a Wolfgang DORDA^a

^a*Section of Medical Information and Retrieval Systems,*

Core Unit for Medical Statistics, and Informatics, Medical University of Vienna

^b*Institute for Scientific Computing, University of Vienna*

Abstract. Lifelong electronic health records can supply valuable information for research, quality management and health politics in addition to supporting treatment of patients. Based on experiences with scientific data analysis in a university hospital environment, requirements on cross-institutional analysis of electronic health records in a healthcare system are discussed. The concept of archetypes can play a key role in this context. Archetypes can be utilized in data analysis for visualization, semantic linkage and finally for standardized data transfer.

Keywords. Archival-repository systems for medical records-EPR-CPR-EMR, Standards, Data analysis-extraction tools, Epidemiological research Hospital IS

Introduction

Introducing standardized electronic health records (EHR) is a strategic e-Health goal in Europe [1] and therefore in Austria as well [2]. Implementing this important project is going to influence *documentation*, *communication* and *analysis* of patient-related data. Services based on standardized *electronic health record architectures (EHRAs)* [3] are intended not only to improve treatment of individual patients but also to give a fresh impetus for research, quality management and health politics.

Data analysis may concern the EHRs of individual patients or may be conducted across patients. The former case is most likely to involve a physician or patient screening an EHR for relevant data in the context of medical treatment. The latter case will normally involve a variety of EHRs analyzed by an investigator or statistician in an effort to identify commonalities or differences.

The focus of the present article is to provide an analysis of requirements that a system for cross-patient and cross-institutional analysis should meet in a heterogeneous distributed healthcare system. Discussing all analysis requirements in detail is beyond the scope of this communication. Therefore we shall use what we consider to be the most important requirements for *local data analysis* in a *hospital environment* (as state

¹ Corresponding Author: Walter Gall, Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, Austria; E-mail: walter.gall@meduniwien.ac.at

of the art) as departure point and focus our analysis on the additional requirements needed for cross-institutional analysis within a *healthcare system environment*.

1. Methods

As a departure point, we take a look at data analyses in a hospital. Alongside scientific studies and requirement compilations such as [4] or [5], we mainly rely on the ArchiMed scientific retrieval system [6] for this purpose. This system allows heterogeneous medical data from different routine applications to be analyzed across patients together with data of clinical studies. The system has been used at the Austrian university clinics in Vienna and Graz for 10 years. Around 3000 data analyses are performed every year.

The goal of our work is to specify requirements for a retrieval system in the healthcare environment, taking into account the recommendations of the Austrian EHR feasibility study [7].

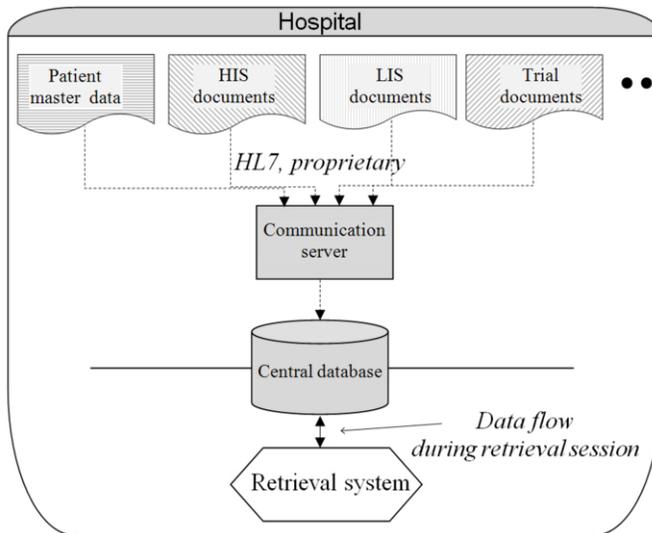


Figure 1. Data analysis in a hospital environment

Figure 1 illustrates a data retrieval system in a typical hospital environment. An ideal system should meet the following important *requirements*:

- (1) **Data integration** by adopting data from documentation systems in a centralized data base that allows data to be analyzed based on a generic data model [8].
- (2) **Support of the three typical retrieval steps** for clinical data, characterized by three sequential activities: *cohort formation* (searching for patients with

specific findings), *selection of variables* within the selected cohort (selecting the patient data of interest and processing them to form a statistical matrix) and *statistical analysis* (ranging from preparatory statistical functions up to utilizing the complete functionality of a statistical analysis package) [6]. The following requirements (3) and (4) are main features of the retrieval steps *cohort formation* and *selection of variables*.

- (3) **Analysis of the time course** of a disease [9] by providing powerful temporal relations for selection conditions (e.g. “within 2 to 4 months after the surgery“).
- (4) **Text analysis** by using text operators for free text analysis and code systems such as ICD as thesaurus support. Additional algorithms are sometimes used to split up word compounds and to build textual roots [10] although the task of creating specialized textual knowledge bases has turned out to take an inordinate amount of time and effort.
- (5) **Support of a large variety of user groups** by the bandwidth of the retrieval functionality, ranging from powerful components for flexible ad-hoc specifications of individual data analyses performed by power users down to prefabricated adaptable data analyses for novices.
- (6) **Intuitive query formulation** by the use of a graphic query editor. This alternative to conventional line-based query formulation is routinely demanded but its implementation has usually been confined to prototypes.
- (7) **Web-based analysis** through a Web interface to execute prefabricated data analyses. This functionality is not absolutely required within hospitals but is frequently used in the context of multicenter studies.
- (8) **Data privacy** ensured by a comprehensive user and authorization system and anonymisation mechanisms for data export functions.

2. Results

Data retrieval performed in heterogeneous hospital environments that have developed over numerous years have a lot in common with data retrieval performed inside a nationwide healthcare system. However, the requirement profile cannot be directly transferred from one to the other. A comparison between Figure 1 and 2 shows that even the very mechanisms of data communication involved will differ greatly. In hospitals, data analyses are usually performed through a centralized database. In a nationwide EHR system, by contrast, documents need to be accessed that are spread over various places.

The following discussion focuses on the adaptations and extensions concerning the above-described requirements for a retrieval system in a nationwide EHR system environment.

- (1) **Data integration.** Communication with the distributed repositories (assuming a structure such as IHE-XDS [11]) should be based on *archetypes* based on an *EHR information model* [12]. This will ensure *semantic interoperability*, which is a fundamental requirement for semantically correct data analysis.
- (2) **Support of three typical analysis steps.** The extended information typology provided by the EHR has to be used. For *cohort formation* and *selection of*

variables, organizational (e.g. type of healthcare provider, federal state) and context-related (e.g. archetype) selection attributes should be available in addition to patient-related ones (e.g. patient, hospital stay). The requirements for statistical analysis remain unchanged.

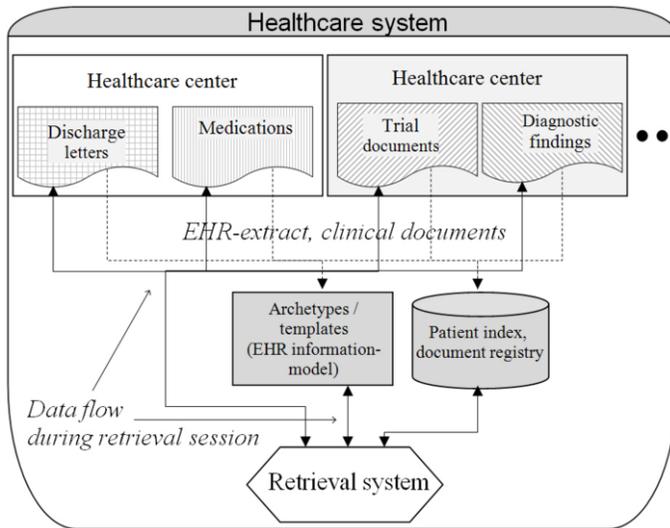


Figure 2. Data analysis in a countrywide EHR system environment

- (3) **Analysis of the time course.** In the EHR the temporal logic has to be combined more closely with components pertaining to content (context-related archetypes). The question arises whether or not a specific temporal sequence can be evaluated jointly when the individual values were collected under different conditions, for example routine examination versus ergometry.
- (4) **Text analysis.** Heterogeneity of data is increased due to the variety of data generators involved (e.g. physicians in private practice, university hospitals, laboratories). Analyses are not only focused on text parameters such as spelling or abbreviations but need to take into account the broader context of data compared to hospital environments. The use of *ontologies* [13] in conjunction with archetypes can contribute to supporting complex semantic analysis of medical terms.
- (5) **Support of a variety of user groups.** The bandwidth of users does not become considerable larger in a healthcare system environment. However, the geographical extension is such that users can hardly be supported by conventional training and care. The on-line help functions need to cover all needs for assistance. As a result, they have to meet advanced quality requirements.
- (6) **Intuitive query formulation.** Due to the larger complexity of cohort formation and selection of variables, the options of query formulation need to be expanded in proper way. Symbols and metaphors are gaining importance.

The EHR-based query languages that are currently being developed [14] need to be implemented in graphic format. Visualizing archetypes can both help to obtain a better overview of the context into which information is embedded and can serve to select variables.

- (7) **Web-based data analysis.** Web applications are unavoidable in a nationwide EHR system environment. Technical challenges notably include the support of query formulations (e.g. by graphic means) based on Web technologies and performance requirements for accessing distributed data. In the communication domain, new possibilities arise with the introduction of asynchronous methods (e.g. Ajax).
- (8) **Data privacy.** Whenever patient data leave a healthcare facility, data protection is a major concern. Investigators using such data inside a hospital frequently do so implicitly in a treatment context that will justify the analysis of data related to individuals. This context is usually lost once the data are accessed in a healthcare system environment. The data can only be analysed in an anonymized fashion. If identifiers are needed for complex multi-step queries (cohort formation, selection of variables), the requirements of *k-anonymity* must be met in addition to *anonymization* and *pseudonymization*. The requirements of *k-anonymity* are met if a cohort encompasses at least *k* data records with the same secondary identification characteristics [15]. When data groups are extracted, care must be taken to avoid any inadvertent inclusion of secondary identification characteristics that might not be consistent with the described criteria of *k-anonymity*.

3. Discussion

Introducing a standardized electronic health record is a defined strategic e-health goal in Europe. In the present article, a number of key requirements have been discussed that systems used for cross-patient analysis in a nationwide EHR system environment should meet. Numerous of these aspects will also apply to intra-patient data analysis as conducted by treating physicians or the patients themselves.

Naturally the requirements discussed in this paper can still be extended (e.g. to include result management), refined (e.g. anonymization algorithms) and examined for the time and effort that will likely go into their implementation. Numerous questions remain to be settled also with regard to *data privacy* as the most important aspect. Where should analysis steps be conducted? How can a document be opened to analyze the medical concepts used (archetypes) while ensuring anonymity at the same time?

Semantic integration of data is a quintessential task. *Archetypes* will play a key role in this connection. This concept is promoted by the three most important EHRA sources (CEN, HL7, openEHR). Their use builds a first step to render heterogeneous existing data interoperable [16] and hence analysable.

Selecting variables for analysis from complex data structures can be supported by visualizing archetypes with graphic user interfaces. Furthermore, *semantic links* can be used in conjunction with ontologies to extend a search through one archetype to other archetypes.

For example, if an analysis of data on diabetes mellitus according to patients is performed by searching for specific values of the medical history and of the blood sugar, the archetype “diabetes” could be utilized first to find the variables “medical

history” and “blood sugar”. Furthermore, the search could be extended through the ontology section of the variable “blood sugar” to include an archetype “laboratory chemistry” containing the synonymous variable “blood glucose”.

4. Conclusion

The aspect of data retrieval is frequently a stepchild in comparison with the aspect of data collection. It is not considered until data collection has already been completed. By that time, it is too late to take the requirements for data analysis into account. Even in standards for EHRAs, the “retrieval” aspect frequently falls into the category “future work”. Retrieval-oriented requirements should already be considered when a standardized EHR system is introduced. Only then can this veritable gold mine of information be used efficiently for quality management, epidemiologic research and health politics.

References

- [1] Commission of the European Communities. e-Health - making healthcare better for European citizens: An action plan for a European e-Health Area. http://eur-lex.europa.eu/LexUriServ/site/en/com/2004/com2004_0356en01.pdf (accessed 2008-02-10).
- [2] Austrian Federal Minister for Health and Woman. Excerpt from the 2005 Health Reform Act, Fed. Law Gaz. I no. 179/2004. http://www.bmgfj.gv.at/cms/site/attachments/3/4/1/CH0027/CMS1043931577060/health_care_quality_act.pdf (accessed 2008-02-10).
- [3] Blobel B. Advanced EHR architectures--promises or reality. *Methods Inf Med* 2006;45(1):95-101.
- [4] Safran C, Chute C. Exploration and exploitation of clinical databases. *International Journal of Biomedical Computing*, 1995; 39:151-156.
- [5] Ammenwerth E, Buchauer A, Haux R. A requirements index for information processing in hospitals. *Methods Inf Med*. 2002; 41(4):282-8.
- [6] Gall W, Sachs P, Duftschmid G, Dorda W. A retrieval system for the selection and statistical analysis of clinical data. *Med Inform Internet Med*. 1999; 24(3):201-12.
- [7] Austrian Federal Minister for Health and Woman. Feasibility study ELGA. http://www.bmgfj.gv.at/cms/site/attachments/7/2/0/CH0513/CMS1169796766007/machbarkeitsstudie_elga.pdf (in German, accessed 2008-02-10).
- [8] Dorda W, Wrba T, Duftschmid G, Sachs P, Gall W, Rehnelt C et al. ArchiMed: a medical information and retrieval system. *Methods Inf Med*. 1999; 38(1):16-24.
- [9] Shahar Y, Combi C. Timing is everything. Time-oriented clinical information systems. *West J Med*. 1998; 168(2):105-13.
- [10] Dorda W. Data-screening and retrieval of medical data by the system WAREL. *Methods Inf Med*. 1990; 29(1):3-11.
- [11] IHE ITI Technical Committee. IHE IT Infrastructure Technical Framework, vol. 1 and vol. 2.
- [12] Beale T. Archetypes and the EHR. *Stud Health Technol Inform*. 2003; 96:238-44.
- [13] Smith B, Ceusters W. An ontology-based methodology for the migration of biomedical terminologies to electronic health records. *AMIA Annu Symp Proc*. 2005; 704-8.
- [14] Ma C, Frankel H, Beale T, Heard S. EHR Query Language (EQL) – A Query Language for Archetype-Based Health Records. *Stud Health Technol Inform*. 2007; 129:397-401.
- [15] Sweeney A. k-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 2002; 10(5): 557-570.
- [16] Kalra D, Blobel G. Semantic Interoperability of EHR Systems. *Stud Health Technol Inform*. 2007; 127:231-45.

Consent-based Access to Core EHR Information: the SUMO-project

Vigdis HEIMLY

Norwegian Centre for Health Informatics, Trondheim

Abstract. Lack of access to updated medication information is a challenge for healthcare providers in Norway. Drugs are prescribed from different sources as the patients GP, private specialists, emergency care, hospitals and doctors in the patient's family. In order to provide healthcare providers that need access to the patient's medication information fast, a project for consent-based access to a core-EHR has been established. All major EHR-system vendors in Norway participate in the project that is funded by national health authorities, Innovation Norway¹ and the municipalities. The core-EHR provides a generic basis that can be used as a pilot for a national patient summary.

Keywords. consent, EHR, patient summary, medication

Introduction

It is estimated that medication-related errors cause 190 deaths/year in Norway (population 4.5M) and that 160.000 unintended incidents occur outside hospitals every year. It is assumed that a considerable number of these medication-related errors are due to faulty or inaccessible drug information

When the Directorate of Health and Welfare launched their strategy, Teamwork 2007 [1] in 2003, a national municipality programme was part of the plan. A group of lighthouse projects were initiated and given limited funding from the government as part of this programme. The idea was that these projects should become examples to be followed by other projects, and that necessary national standards should be developed in relation to the projects. The projects would require extensive involvement from the municipalities in terms of both own workload, software development and infrastructure. In order to reduce the number of drug-related unintended incidents, one of the national lighthouse projects had focus on this area. The project was established by the municipality of Trondheim in 2003.

The directorate of health and welfare initiated 5 projects:

1. Consent-based access to core-EHR, drug-information (Trondheim)

¹ Innovation Norway, prior the Norwegian Industrial and Regional Development Fund (SND), helps to provide or arrange financing, link customer enterprises to know-how and help them build networks for their innovation projects.

2. Shared individual plans, continuity of care (Kongsvinger/Eidskog)
 3. Intermunicipal telemedicine (Nord-Gudbrandsdal)
 4. Extended electronic message exchange for better cooperation between hospitals and municipalities (Stavanger)
 5. A model for implementation of EHR in community care (Sandefjord)
 6. Use of secure e-mail within community care, laboratory results and discharge summaries from hospitals to nursing homes (Tromsø)
- The sixth project in Tromsø, was already started before the programme started.

Other related Projects

In parallel with the lighthouse-project, the ELIN-k project was established by the Norwegian Nursing Association and KS². The main goal of the ELIN-k project is to ensure that healthworkers in community care can have access to updated health information from other GPs and hospitals. Information is transferred by standardised electronic messages.

The ELIN-k project uses the same projectmodel as the ELIN-project that was initiated by the Norwegian Medical Association. The main goal for the ELIN-project was to ensure that the GPs got solutions for exchange of medical information that were integrated into their electronic health record (EHR)-systems and supported their daily work routines. A requirement for the solution was that national standards should be supported. Innovation Norway has partly funded the vendors of EHR-systems work with implementation of the needed standards. The standards have been developed by the Norwegian Centre for Health Informatics (KITH) based on the users requirements from the ELIN-project. As a result, all the vendors have implemented standards for communication of referrals, laboratory results and discharge summaries in their systems. The implementations have also been tested by the national test- and approval service operated by KITH.

1. The lighthouse-project in Trondheim

The project focuses on a group of mostly elderly people (age 80+) living at home but with an extensive need for healthcare services. In order to provide the healthcare providers from different organisations with updated information about the patient's medication, it was decided to develop a system based on core-EHR [2].

The core EHR will be updated automatically when pharmacists, hospitals and other healthcare providers participating in the projects send information to the GP. The patient's GP is responsible for the core EHR and will check all information received from other healthcare providers in order to detect any medication inconsistencies

The patient and his GP jointly decide what information other healthcare providers may need in different situations based on the patients consent. The patient may at any time change or withdraw his consent, and change the distribution rules.

The patient and his GP jointly decide the distribution rules to be followed:

- Distribution rules are partly based upon ENV13606-2

² The Norwegian Association of local and regional authorities (KS) is a national organisation for municipalities, counties and public enterprises under municipal or county ownership,

- Who may receive healthcare information
 - Explicitly named healthcare professionals
 - Healthcare professionals having a specific role
- Why may they receive healthcare information
 - A set of standardised requests is used to specify why the requesters need healthcare information
- When may they receive healthcare information
 - The patient's consent is either given for a specified period of time or for a specific period of care

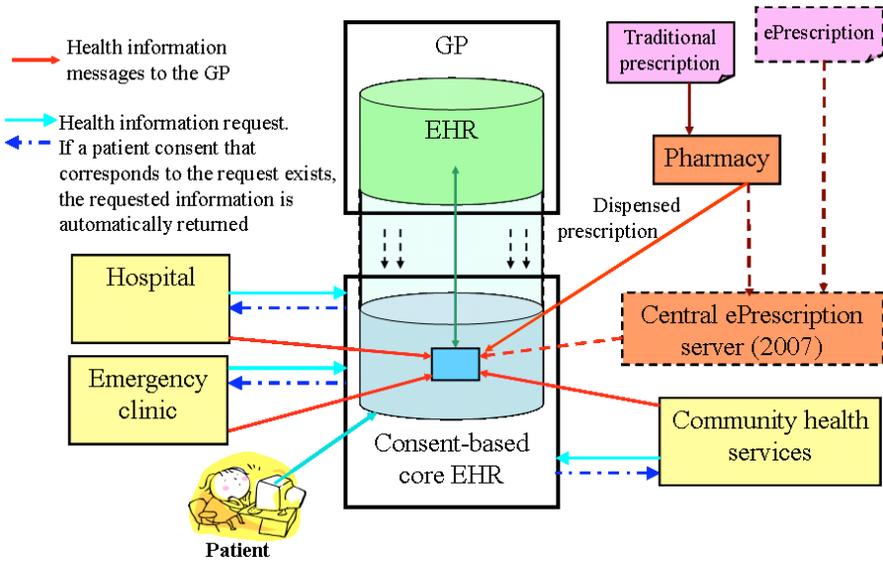


Figure 1. Core-EHR

To comply with Norwegian legislation, the core-EHR is updated by messaging. Access to the core-EHR can also be provided by web-services

The selected information is automatically copied from the GP's EHR to a "core EHR" on separate server that handles the requests. A server-based solution has been chosen in order to keep the core-EHR accessible outside the GP's regular workhours. The core-EHR is run on a server in the Trondheim municipalities net, and access from the Norwegian Health Net to the core-EHR will also be established.

The project uses the EHR kept by the patient's regular GP as the main source of information. All Norwegian healthcare providers are obliged to give the patient's regular GP relevant health information unless the patient explicitly refuses.

Other actors who need and are will receive updated drug-information based on requests sent to the GP's EHR-system. Each request is automatically checked against a set of distribution rules and if the request is covered by one of these rules, a message containing the requested information is automatically returned

1.1. The EHR-systems that need to communicate with the core-EHR in the Trondheim region

The hospitals in Norway are organised within 5 regional health authorities, owned by the government, and Trondheim municipality mainly communicated with hospitals that are located within the Regional Health Authority (RHA) of Mid-Norway. All the hospitals in RHA Mid-Norway are using the same EHR-system from Siemens. The EHR-systems for the hospitals are all operated at the same location, but the hospitals databases are separate due to Norwegian legislation, so although it would technically be possible to share EHR-information between the hospitals, this is not possible without the patient consent.

The municipality in Trondheim has installed an EHR-system in community care from the vendor Tieto Enator. This system is in use at nursing homes and for homecare. The system is used by both for administrative workers and healthcare providers.

The GPs in Trondheim have been using EHR-systems for more than a decade. They do not keep paper records any more, and rely solemnly on the electronic record. The GPs use the EHR-system actively while the patient is in the office, and they are very concerned by the fact that the EHR-system must be easy in use, and that new functionality must be added in a way that does not force them change work routines to something that is less efficient than today. There are currently EHR-systems from two vendors, Profdoc and Hove in use within the municipality.

1.2. Communication between the EHR-systems before the core-HER was established in Trondheim

By the time the projects started, there were limited possibilities for communication between the EHR-systems. The GPs received laboratory results and discharge summaries and had electronic communication with NAV, the Norwegian Labour and Welfare Organisation. A pilot for sending electronic referrals from GPs to hospitals and specialists was also established.

1.3. Infrastructure for the core-EHR

The hospitals, GPs and the caretakers in the municipality of Trondheim all have access to the Norwegian Health Net. The Norwegian Health net is owned by the RHAs, and has so far mainly been used by the hospitals, but the number of GPs connected to the net is rising. In November 2007, 15% of the Norwegian municipalities are connected to the Norwegian Health net

2. From lighthouse project in Trondheim to SUMO

Other municipalities in Norway soon also showed interest for the core-EHR, but they had started out at a slightly different angle than the project in Trondheim. Stavanger municipality wanted to use regular messaging between the actors to keep the drug information updated, and the Norwegian Centre for Telemedicine in Tromsø had also initiated a medical card project together with Tromsø municipality based on use of web-services.

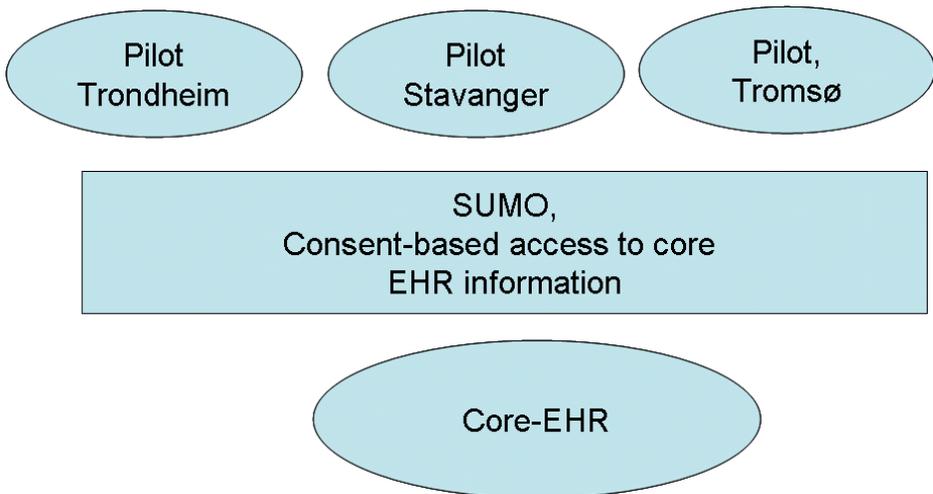


Figure 2. SUMO

It was soon evident that the three projects would benefit from cooperation. Development of common standards and implementation of the same standards for all EHR-vendors would be highly advisable. The project in Trondheim municipality did only count for 1/3 of the EHR-vendors in Norway, but all major vendors would be covered in a joint project with Stavanger, Tromsø and Trondheim. The three projects thus decided on cooperation, and the projects in Trondheim and Tromsø are now using the same core-EHR system, and all three projects use the same set of standards. To comply with Norwegian legislation, the core-EHR is updated by messaging. Access to the core-EHR can also be provided by web-services.

An application for funding of the EHR-system vendors development of client-modules in their EHR-systems was sent to Innovation Norway, and was approved late 2006. The same project model as for the ELIN-project was again used. A project to coordinate the vendors work, called SUMO, was established in early 2007.

The SUMO-project has been running a number of Workshops with vendors and users throughout 2007. Messaging standards for administration of the core-HER, and exchange of EHR-information, have been developed by KITH based on requirements from users and vendors in the SUMO-project. The new messaging standard for EHR is based on reusable components [3], and will probably prove to be a useful basis for development of future communication standards. Requirement specifications for the client-modules in the vendors EHR-systems have also been developed. The vendors are now at a stage where they implement the standards and the client-modules in the EHR-systems. The project is due to finish in June 2009, and the first pilots will be up and running in spring 2008. The core-EHR is already developed, and is being tested by Trondheim and Tromsø municipality.

The project works closely together with the ELIN-k project, the national ePrescription project run by the Directorate of Health and Welfare and the Norwegian pharmacies organisation. It is crucial for the vendors that larger national projects are coordinated and that the same set of standards can be shared and reused across projects.

As an example will the structure of the medication in the SUMO-project be the same as defined in the ePrescription project.

3. From SUMO to national core-EHR and patient summary

It is a need for a core-EHR solutions that can be used for more purposes than drug information. Examples are: Shared individual plans to support continuity of care, summary of the patient's contacts with health providers in different organisations, and core-EHR information as important diagnoses, allergies and contact information. The core-EHR and the model from the SUMO-project can provide a good basis for a more general Norwegian core-EHR. The RHA's ICT-organisation, NIKT, has now initiated a national project to evaluate how a core-EHR can be realised in the coming years, and the suggested solution for SUMO is likely to become a model for future work in this field in Norway.

References

- [1] Teamwork 2007, Directorate of Health and Welfare, 2007
- [2] Løsningskisse fyrtårn Trondheim: Legemiddelopplysninger i Samtykkebasert kjernejournal, Centre for Health Informatics, KITH-rapport 29/05.
- [3] Kommunikasjon av EPJ-innhold i SUMO-prosjektet, Metodebeskrivelse, Norwegian Centre for Health Informatics, KITH-rapport 08/07.

Facilitating the openEHR approach – organizational structures for defining high– quality archetypes

Christian Dominik KOHL^a, Sebastian GARDE^b, Petra KNAUP^a

^a *Institute for Medical Biometry and Informatics, University of Heidelberg, Germany*

^b *Central Queensland University, Austin Health, Heidelberg VIC, Australia*

Abstract. Using openEHR archetypes to establish an electronic patient record promises rapid development and system interoperability by using or adopting existing archetypes. However, internationally accepted, high quality archetypes which enable a comprehensive semantic interoperability require adequate development and maintenance processes. Therefore, structures have to be created involving different health professions. In the following we present a model which facilitates and governs distributed but cooperative development and adoption of archetypes by different professionals including peer reviews. Our model consists of a hierarchical structure of professional committees and descriptions of the archetype development process considering these different committees.

Keywords. Computerized medical record systems openEHR archetypes, domain knowledge governance, modeling, quality management.

Introduction

The openEHR architecture for Electronic Health Records (EHRs) is a two level modeling approach based on ‘archetypes’ [1]. Archetypes ‘are reusable, structured models of clinical information concepts that appear in EHRs ...’ [2]. Using openEHR [3, 4] means using existing and creating new archetypes. It is a central idea of the openEHR approach to use one archetype in as many systems as possible to reduce time and effort for design and to enable semantic interoperability among different systems [5]. However, building such high-quality archetypes, which are accepted among the particular domain experts, requires appropriate governed development processes [2].

Three prerequisites for reusable, high-quality archetypes

To draw an analogy to other development processes, e.g. for standards or evidence based medicine, three conditions should be fulfilled for development or adoption of an archetype, which are independent from the intended area of application – either for a single hospital, multiple institutions or a complete country:

- The design / modification of an archetype should consider evidence based knowledge and design recommendations.

- The design/modification of an archetype should be peer reviewed. Both design and review should be conducted in a flexible framework.
- Released archetypes should be published in an open-access repository and evaluated in terms of their usability and professional correctness.

The aim of our paper is to analyze the structures and processes that are necessary to assure these three prerequisites. Furthermore we present a model which contains the requested structures and processes as well as their application to enable the international and inter-professional coordination of archetype development and maintenance.

1. Material and Methods

The creation of our model is a design research approach [6]: We systematically combined field reports such as [7] with expert opinions and own experiences. Furthermore we considered experiences from other standards development processes and identified the special requirements of the archetype development process. We consequently carried on the work of Garde et al. who gathered ‘requirements for the development of an archetype repository which provides comprehensive support for archetype development cycles ...’ [8]. They conducted workshops on archetype development for clinicians, for their requirement analysis.

2. Results

Basic components of our model are descriptions of necessary structures including a role concept and process descriptions based on these structures.

2.1. Necessary structures

1. Users / user group: Everybody who does or will use a system based on archetypes and everybody who designs or modifies archetypes is called a user. The single users have different professional backgrounds: Physicians, nurses, computer scientists... All users together form the ‘user group’.

2. (Professional) Committees: To facilitate peer review, a hierarchical tree structure of committees has to be established. Every committee represents a particular division of the health domain. Committees closer to the root are more general whereas ‘leaf committees’ are highly specialized. The root node is built by the clinical review board (Figure 1).

A user from the user group can join one or more committees according to his or her professional background and interest. For instance a nurse in an emergency department might join the ‘emergency nursing group’ in the ‘nursing’ branch but also the ‘emergency cooperative care’ group in the ‘cooperative care’ branch.

Every committee (including the clinical review board) has a chair and is responsible for a certain set of archetypes.

3. Clinical review board: The clinical review board is a group of experts with different professional backgrounds representing the different subgroups of the user group including government representatives where appropriate.

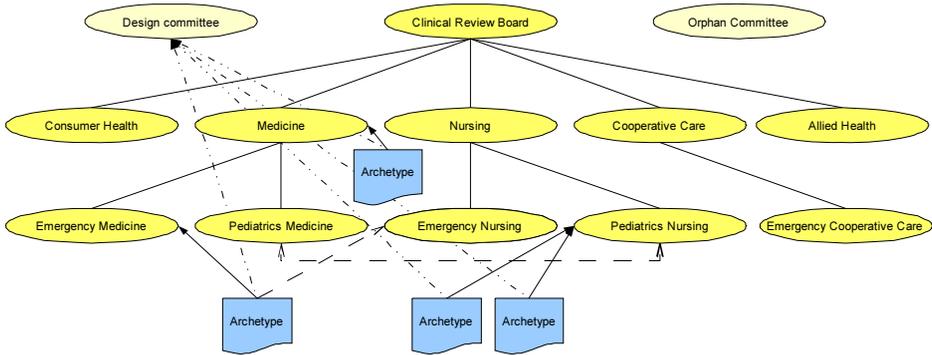


Figure 1: Possible committee structure

4. Design committee: The design committee observes that archetypes are sensibly developed and modified from a design point of view. Furthermore, the committee offers all users a ‘technical counseling’. It assembles and provides information material about relevant standards and programming recommendations.

The chair of the design committee is also member of the clinical review board to assist it in design and technical issues. Some members of the design committee could be representatives of companies developing openEHR related technology.

5. Orphan committee: This committee takes care of archetypes which at the moment are not assigned to a certain professional committee.

2.2. Processes descriptions

Let’s assume someone needs an archetype for a certain problem: He or she first searches the public repository to find a suitable archetype, which works like a typical repository for cooperative software development (e.g. a subversion repository). In the long run, metadata stored in an archetype OWL ontology [2, 9] should ease that search.

If a suitable archetype already exists, it can be directly used. If not, a new archetype has to be designed or, if there has been an advance in medical knowledge, an existing one has to be adopted.

Usually several archetypes are necessary to establish a certain system: Every required archetype has to pass the described processes.

2.2.1. Designing a new archetype

1. Choice of committee: The user who requires a new archetype has to pass a short description of its intended use to that committee which should be responsible according to his judgment.

If no committee seems suitable, the user may pass the request to the orphan committee, which then assigns the request to a committee or decides to establish a new committee respectively.

A user who needs a certain archetype is one possible starting point of this process; another could be that a certain committee sees the need of a special archetype.

2. Allocation of an archetype: The requested committee has now to decide whether it is responsible for the archetype or not. If not, it may pass the request to another appropriate committee (which now has to debate its responsibility) or to the Orphan Committee which may assign it to a certain committee that has to accept the assignment.

3. Review of request: After declaring the responsibility for a request, the particular committee decides whether to accept the request (which means the request seems to be reasonable to the responsible committee) or not. If the committee declines the request, it informs the enquiring person about the reasons so that he or she is able to modify the request.

Once an archetype is accepted, the responsible committee has to complete its description in the metadata.

4. Linkage of further committees / observers: There may be an additional committee interested, so that a link can be set. This takes place either by request of the responsible committee, the 'second' committee or the Orphan Committee.

We distinguish two kinds of links for reviewing: 'Minor review links' and 'major review links' (cf. step six). Additionally, the archetype is assigned to a member of the design committee.

Furthermore any interested user can link himself to the archetype and becomes an 'observer'. Whenever an 'observed' archetype is altered, a linked user gets an email notification containing a summary of the modifications. When users can also link themselves to a certain committee they automatically become observers to all archetypes linked to that committee.

5. Development and supervision: Every archetype is developed under the supervision of the responsive committee (certain committee members might be responsible for certain archetypes) and the supervision of the assigned member of the Design Committee. The technical development itself might be done by the user who initiated the request and / or by a member of the responsible committee and / or by a third person.

6. Review of archetype (responsible and linked committees): When the persons involved in the development of an archetype finished the archetype, they set the state of the archetype from 'author draft' to 'committee draft'. That is a signal to all committee members to review it and to discuss problems and alternatives. The review of an archetype includes the generation of a validity report and compatibility checks. That applies for all committees, which are 'major review'-linked to the archetype. Both, major and minor review links may be associated either with a particular archetype (= 'archetype links') or with a particular committee (= 'committee links'). Thus, every archetype is automatically linked to all 'committee linked' committees.

If the responsible and all assigned committees agree that there should be certain changes, the state of the archetype is set back to 'author draft' so that the developers can implement the suggestions and start another review.

If there are no more open issues, the archetype-state is set to 'committee-final' and the archetype is passed to the committees 'minor review' linked to it. They have to come to an agreement whether to accept the proposed archetype or to suggest modifications. In the second case the state of the archetype is set back to 'author draft' for revision and step six starts over.

If all linked committees as well as the primarily responsible committee accept the archetype, it is passed to the superior committee.

If a consensus between the responsible committee and one or more linked committees is not possible, the clinical review board can be called to mediate.

7. Review of an archetype (superior committees): The superior committee discusses the archetype and either accepts it and passes it on to its superior committee or, the state of the archetype is set to ‘author draft’ and the archetype is sent back to the primarily responsible committee for modification. Afterwards the archetype starts again his way up to the root node. With the notification that there is an archetype to review every committee gets a short summary of the changes since the last version.

If a superior committee considers a requested change as a ‘minor change’, it is modified by the developers and reviewed by the responsible committee without starting the complete way up again. It continues its way at the superior committee which requested the minor change.

If an archetype is reviewed by a committee which is not responsible for that archetype, none of the committees which have a committee link to the current committee gets the archetype for review.

8. Publication of an archetype: As soon as the archetype is accepted by the clinical review board, its state is set to ‘final’ and it is transferred into the public repository and accessible by everyone.

To ease the selection of a set of archetypes for a certain purpose from the public repository, a professional committee might compose a set of archetypes suitable for its professional field.

2.2.2. Adoption of an existing archetype

The process of adopting an existing archetype is in line with the design of a new archetype except of step one and two: As an existing archetype is already assigned to a committee, so that committee has just to decide if it accepts the request.

3. Discussion

The proposed theoretical model describes processes and structures to facilitate a cooperative development and extension of openEHR-archetypes including peer review and design control in the large. This should lead to high-quality archetypes. Furthermore, our model is in principally applicable to the design of other (non-archetype) knowledge resources – for instance HL7 CDA [10]. Archetypes, however, offer the advantage of being formal specifications and being intuitively understandable for clinicians.

Having a common, international repository of archetypes, as the model implies, could reduce time and effort for designing archetypes and could advance interoperability among different systems [11]. If introduced at an international level the model could for instance avoid the design of two incompatible archetypes for the same concept. Thus, the model could support fundamental ideas of openEHR.

Nevertheless, the proposed model is a theoretical model, based on an idealized starting point. Therefore, a practical realization is necessary to check its applicability and to optimize the model: As the model contains cycles, infinite loops within the design process might occur. Furthermore, if someone has to review the same archetype a few times, the review quality might decrease. Rules should be established to avoid

structural and quality problems - for instance a limit of the maximum number of development-and-review iterations.

4. Conclusion and perspective

A cooperative and distributed development and extension of archetypes is necessary to reduce time for archetype designing and to improve semantic interoperability, but not easy to perform. The proposed model could help to realize such a scenario. To facilitate and coordinate the necessary communication among the committees and to integrate the archetype development process, simple ‘wikis’ and content versioning systems are not sufficient.

As a first step sufficient tools supporting the described processes and structures have to be implemented and evaluated in a smaller context. Once tools and model are well engineered and evaluated, both should be introduced in larger (international) context.

References

- [1] L. Michelsen, S.S. Pedersen, H.B. Tilma, S.K. Andersen, Comparing different approaches to two-level modelling of electronic health records, *Studies in Health Technology Informatics*, **116** (2005), 113–118.
- [2] S. Garde, P. Knaup, E.J.S. Hovenga, S. Heard, Towards Semantic Interoperability for Electronic Health Records: Domain Knowledge Governance for openEHR Archetypes, *Methods of Information in Medicine*, **46** (2007), 332–343.
- [3] Introducing openEHR
(http://svn.openehr.org/specification/TRUNK/publishing/openEHR/introducing_openEHR.pdf – (accessed 20 September 2007))
- [4] D. Kalra, T. Beale, S. Heard, The openEHR Foundation, *Studies in Health Technology Informatics*, **115** (2005), 153–173.
- [5] B.G. Blobel, K. Engel, P. Pharow, HL7 Version 3 compared to advanced architecture standards, *Methods of Information in Medicine*, **45** (2006), 343–353.
- [6] J. Buck, *Entwurf einer umfassenden elektronischen Patientenakte für die Neonatologie der Universitätsklinik Heidelberg unter Verwendung des openEHR-Ansatzes*, Thesis, Heidelberg, 2006.
- [7] Design Research in Information Systems. V. Vaishnavi, B. Kuechler Section Editors. 2007 (<http://www.isworld.org/Researchdesign/drislSworld.htm> - (accessed 10 October 2007).
- [8] S. Garde, E.J.S. Hovenga, J. Gränz, S. Foozonkhah, S. Heard, Managing Archetypes for Sustainable and Semantically Interoperable Electronic Health Records, *electronic Journal of Health Informatics*, **2** (2007), e9.
- [9] *OWL Web Ontology: W3C Recommendation*. M. Dean, G. Schreiber Editors. W3C, 2004. (<http://www.w3.org/TR/2004/REC-owl-ref-20040210/> - (accessed 27 September 2007).
- [10] R.H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F.M. Behlen, P.V. Biron, A. Shabo Shvo, HL7 Clinical Document Architecture: Release 2, *Journal of the American Medical Informatics Association*, **13** (2006), 30–39.
- [11] R. Chen, G. Enberg, G.O. Klein, Julius: a template based supplementary electronic health record system, *BMC medical informatics and decision making*, **7** (2007), 10.

Implementation of an electronic medication system and disregarded power of the record

Henriette MABECK¹

Danish Institute for Health Services Research, Copenhagen

Abstract. Though hospitals managers wish to have a unified medication system, physicians continue double registration of prescriptions. The traditions of prescribing both electronically and on paper, as well as the power of the medical record are some of the disregarded elements when introducing a computerized physician order entry system (CPOE). The result is that the explicit goal to eliminate double registrations is not fulfilled. The pattern of prescriptions continues unchanged and CPOE replaces the abandoned paper based medication card. In this case study a symmetric approach reveals how double registrations on one hand are considered a safety problem, but on the other hand double registrations intercept and prevent medication errors.

Keywords. Computerized physician order entry system, CPOE, medication errors, Actor network, evaluation, organisational change, communication.

Introduction

Unintended consequences and errors in the medication process are frequent iatrogenic injuries [1; 2]. In 2004 the National Board of Health in Denmark received 1803 reports of medication errors [3]. Errors due to transcriptions are varying from 11% to 56% [1-3]. Computerized physician order entry (CPOE) is supposed to be able to eliminate those errors, and that it will imply: "*medication prescriptions are only written in one place, and all the information about the patient's medicine will be accessible in one version. This will minimize the medication errors*"[4]². There is a general agreement that double registration is an important factor in medical errors, and that it can be reduced by CPOE [2; 3; 5; 6]. This paper enlightens two issues in relation to the physicians' practise of continuing double registration. Firstly it will analyse why physicians continue entering prescriptions in two places when they are expected to use CPOE only for prescriptions. Secondly it will illustrate how double registrations can be used as quality measure in the clinic, where a patient is treated by different physicians.

¹ Address for correspondences: hm@dsi.dk

² Underlined by author

1. Research methodology

The study is a case study [7; 8] from a Danish University Hospital. The study analyses, how professional traditions and history interact with the use of CPOE, and how CPOE interact with the medical record, when the combination of paper and computer characterises the daily practise. In this study physicians used a traditional paper based record together with a CPOE which is only used for prescription, dispensing and administration of medication.

Ethnographic methods as observations and interviews were used in this qualitative study [9; 10]. The data collection has moved from an open description to a more selected and focused approach [9; 10]. In the first period the data collection was inspired from grounded theory [11; 12]. Later it was more focused on how physicians use and combine CPOE with different artefacts in their daily work.

Data collection has been conducted in three steps: Observations, data analysis and semi-structured interviews with physicians and nurses. The observation guide was grounded in ethnographic questions such as “What is the situation?” , “Who participates?”, “What do they do?”, “Where does the act take place?”, “When do they use CPOE?”, “When don’t they use CPOE?”, “What do they use besides CPOE?”. 28 semi-structured interviews were done in 250 hours observations time. An interview guide had been created on basis of the selected themes from the observations [13].

2. Results

Notification of prescriptions has traditionally a twofold purpose to the physicians. The prescription primarily helps the physician’s memory, and act as information source for other physicians involved in the treatment plan. Secondly, the prescription serves as information and guidance for the nurses. Before CPOE implementation physicians used to write prescriptions on a medication card. In addition to this card, they used to dictate prescriptions in the medical record. After the CPOE introduction physicians have kept their practise to record prescriptions in two different and separate systems with two different purposes. Occasionally they include fewer details into the prescription, but apart from this the prescription process is left unaffected by the CPOE.

The question is why the practise has not changed in direction of unified medication, although this had been defined as an explicit goal for the CPOE implementation? One physician explains: *“sometimes it is difficult to find out the reason for the prescription, if you don’t have the prescriptions in the record. Therefore, I always enter prescription in the record. I guess it is because you have to explain, why you do as you do”*. Another physician says: *“by dictating prescriptions in the record, I can see the medication when I see the patient again”*. This quote implies he gets his information about prescriptions from the record and not from CPOE. The physicians tells that prescriptions in the CPOE officially are prevailing. Observations and interviews show that in practise, it is prescriptions in the record which are prevailing in the eyes of the physicians. The physicians argue they need to have coherence between the patient case and prescriptions, and state: *“You have to have coherence to the history, so everybody can follow, what has happened to the patient”*.

Another reason for double prescription might be more tacit. It is the possibility to discover an error, when there is a discrepancy between record and CPOE prescriptions. During field work the author witnessed several situations of this kind. In most of these

cases the prescriptions in the paper record were correct and the prescriptions in the CPOE systems.

Examples from observations

A: Observation note from ward round

The physician on ward round is reading the medical record. He observes there are five prescriptions in the record, but in the CPOE system there are four prescriptions.

B: Observation note from ward round

The physician on ward round is reading the medical record. He observes medication X is prescribed in the record, but in the CPOE system it is not listed as current medicine. A continuation note some days ago specifies how medication X is to be paused for one day and continued the day after. The clarifying remark in the record missing in the CPOE system where it is discontinued but not resumed the day after as scheduled. The physician discovers the deviation error when reading the note in the record.

C: Observation note from ward round

The physician on ward round is reading the medical record. He discovers nausea treatment is not prescribed as a 'standing order' in CPOE. In CPOE it is prescribed as 'medication on demand'. The normal procedure for nausea treatment in connection with chemotherapy is that the first dose is 'on demand', and the following doses are 'standing orders'. The correction from 'on demand' to 'standing order' was forgotten in this case.

In all three examples the physician becomes aware of a medication error by discovering a discrepancy between the record and CPOE. This happens when the physician is reading the patient record and not when he is looking into the CPOE system. In the first two examples medication is registered in the record but missed in the CPOE. In example B the record note clarifies the prescription plan. This information is not in the CPOE. In example C the physician observes that the patient receives chemotherapy and should have nausea treatment as standing order and not as medication on demand as prescribed in CPOE. These three examples illustrate how physicians discover medication errors, when reading the patient history in the medical record and comparing the record information with the current prescriptions in the CPOE.

3. Discussion

This study has found that despite the explicit goal to unify medication prescriptions the physicians have not changed their way to carry out prescriptions in parallel (paper and CPOE). There are two major reasons why physicians have not changed this habit:

- Missing compliance between the linear computerized order entry process on the one hand, and a non-linear, paper-based, even messy workflow in the clinical environment
- The power hidden in artefacts and local value, routines, and conventions.

The missing compliance is due to the designers' need to simplify workflow complexity [14]. Figure 1 illustrates how designers often comprehend the medication process. In their eyes medication is a direct single stringed process [4]. There is no relation between the prescription and other patient-related information. The prescription process is isolated from overall patient assessment and diagnosis. This process model goes directly from physician to nurse to the patient.

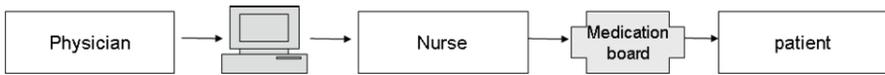


Figure 1 Linear and expected medication process with CPOE

In practice prescription is not an isolated process. In medical practice prescription is not unambiguous. It is divided in two, and it is an element in a long sequence of action patterns. For a physician prescriptions have different meanings and functions depending on the situation and context [14-17]. The prescription has a double purpose. It passes information to the nurses, and it also provides information to the prescribing physician and his colleagues. In Figure 2 the real medication process is illustrated in a simplified model. At ward round the physician and the staff nurse discuss the patient case and make a treatment plan. Then the physician prescribes medication in the record *and* in CPOE. The record prescription is information to himself and his colleagues. The CPOE prescription is an one-way information route directed to the nurses. If a prescription needs to be executed the staff nurse gives a message or a note to the nurse, who looks after the patient. He or she will then carry out the order. Figure 2 illustrates how different artefacts and several people are involved in the prescription process.

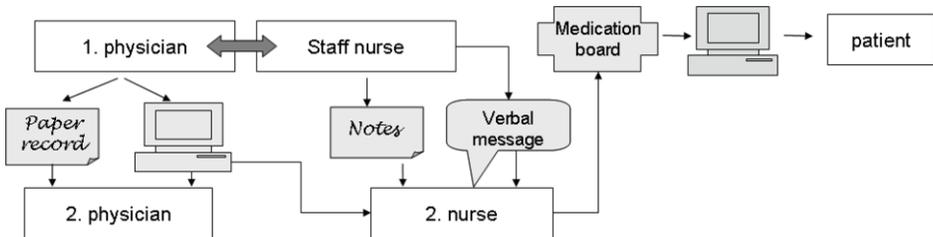


Figure 2 Prescriptions in CPOE – the complex route from physician to patient

The hidden power in the medical record is a result from long traditions. Different physicians cooperate sequentially when treating a patient. The medical record supports the physician memory and act as a communication mediator between the numerous physicians involved in the treatment process [16]. The structure of patient information in the medical record is a working practise, physicians have acquired in generations

[18]. “*You can read from the description, why there was a change and what kind of change.*” one of the physicians explains. Therefore the record has a very high priority in physicians’ mind. This is reinforced by the tradition where the record *belongs* to the physician. The record is where physicians can demonstrate their professional skills to their peers. This power and influence of the record is left out in the idea of CPOE. In an actor network perspective the strength and power of an actor (or actant as Latour calls it) depend on the relations they constitute in a network [19; 20]. This explains why the medical record and prescriptions are strongly related.

The role of the medical record and the local traditions is forgotten or neglected when discussing workflow in relation to CPOE. In the view of designers the record appears as a black box, where there is input and output, whereas in reality the record is shaped in a long process of negotiations and configurations. Viewed as a black box it contains and hides a complex network of traditions, rules, values and power [21]. When CPOE is introduced the hidden power of the record emerges by the way physicians do not change the way they use the record. Changes in prescription patterns are not realized just because the management dictates it. Changes depend on the associations between actors. The association between physicians and the medical record makes the record a powerful artefact. This is one of the more implicit or tacit reasons for continuing the practise of double prescription. The record contains value and gives power to the physicians because it is their property and embraces the whole history of a patient.

A main rationale for CPOE is to eliminate double registrations. In this view double registration is a reason for errors. A more symmetric approach would insist on a balanced view. Examples from this case demonstrate that the prescriptions in CPOE are invalid. This study indicates that double registration is not an unambiguous problem. In different cases double registration may constitute informal and tacit quality control. When physicians compare the record notes and CPOE prescriptions they occasionally discover discrepancies which give rise to further examinations.

This study illustrates how prescriptions in the record are more accurate and up to date than prescriptions in CPOE. Although the CPOE is, officially, the right place for prescriptions, the prescriptions in the record is more reliable. CPOE prescriptions are considered secondary by the physicians and therefore they tend to be less valid. The division of a prescription reflects the relation between the record and the medication card. The record has maintained or even strengthened its position as the prevailing information source for the physicians.

The effort to unify prescriptions has failed because physicians have a deep intrinsic motivation to keep information in the record valid. The record is where physicians demonstrate expertise and get appraised by peers. The CPOE and the previous medication card play a more compulsory role and act as communicating system with nurses. This fundamental difference in the roles and understanding of the record and CPOE is not addressed in the design or implementation of CPOE and this might one of the reasons it hasn’t changed. Even when authorities have decided, that CPOE must be the primary source, prescriptions in the record still are the most valid. Power is not a cause to act – power is a result of acts [20]. The power of the record and the way it interacts in practise is a result of negotiations, traditions and values – the appraisal of colleagues counts more than rules or evaluation from authority.

4. Conclusion

This study shows that double registration helps physicians to discover medical errors, when they compare prescriptions in CPOE with the record. The prescription pattern is continued. The record is the prevailing tool for physicians to communicate to each other and to comprehend and form a patient case. This entails to ensure that the information in the record is complete. Physicians regard CPOE as a tool for nurses and it is not valued as high as the record. Therefore physicians continue to make two prescriptions – one for their peers in the record and one for nurses in the CPOE system. This hidden value and power of the record and the complex prescription process has been neglected in design and implementation of CPOE.

Acknowledgements

Thanks to Århus University Hospital Haematological ward for letting the author follow the nurses' and doctors' daily work and carry out interviews.

Reference

- [1] Benjamin DM. Reducing medication errors and increasing patient safety: case studies in clinical pharmacology. *J Clin.Pharmacol.* 2003; 43(7):768-83.
- [2] Lisby M, Nielsen LP, Mainz J. Errors in the medication process: frequency, type, and potential clinical consequences. *Int.J Qual.Health Care* 2005; 17(1):15-22.
- [3] Sundhedsstyrelsen. Temarapport 2005: Medicinering. København: Sundhedsstyrelsen; 2005.
- [4] Århus Amt. Elektronisk patientjournal i Århus Amt - rapport fra epj-projektgruppen februar 1999 "den blå rapport". Link to PDF: <http://www.aaa.dk/aaa/epjrap99.pdf>. [Retrieved: 10-10-2006].
- [5] Douglas GP, Deula RA, Connor SE. The Lilongwe Central Hosp. Patient Manag. Info. System: a success in comp.-based order entry where one might least expect it. *AMIA Annu. Symp. Proc.* 2003; 833.
- [6] Murff HJ, Kannry J. Physician satisfaction with two order entry systems. *J Am med inform assoc.* 2001; 8(5):499-509.
- [7] Yin RK. Case study research - revised edition. Sage; 1989.
- [8] Maxwell J. Qualitative research design - an interactive approach. California: Sage; 1996.
- [9] Spradley JP. Participant observation. USA: Wadsworth; 1980.
- [10] Emmerson RM, Fretz RI, Shaw L. Writing ethnographic fieldnotes. Chicago: Chicago Press; 1995.
- [11] Strauss A, Corbin J. Basics of qualitative research. 2. ed. California: Sage; 1990.
- [12] Glaser BG, Strauss AL. The discovery of grounded theory : strategies for qualitative research. renewed 1995 ed. Aldine transaction; 2006.
- [13] Kvale S. Interview. Hans Reitzels forlag; 2000.
- [14] Berg M, Goorman E. The contextual nature of med. infor. *Int.J Med.Inform.* 1999; 56(1-3):51-60.
- [15] Mol A. The body multiple: ontology in medical practice. 2. ed. Duke University Press; 2005.
- [16] Berg M, Harterink P. Embodying the patient: Records and bodies in early 20th-century US medical practice. *Body & Society* 2004; 10 (2-3):13-41.
- [17] Berg M, Bowker G. The multiple bodies of the medical record: toward a sociology of an artifact. *The sociological Quarterly* 1997; 38 number 3:513-37.
- [18] Berg M. Medical work and the computerbased patient record: A sociological perspective. *Methods Inf.Med.* 1998; 37:294-301.
- [19] Latour B. Technology is society made durable. In: Law J (ed.). *A Sociology of Monsters: Essay on Power, Technology and Domination.* London: Routledge; 1991: 103-31.
- [20] Latour B. *The Powers of association.* In: Law J (ed.). *Power, action and belief - a new sociology of knowledge?* 1986: 264.
- [21] Latour B. *Science in Action.* Cambridge, Massachusetts: Harvard University Press; 1987.

cyberMarathon – increasing physical activity using health-enabling technologies

Maik PLISCHKE^{1,a}, Michael MARSCHOLLEK^a, Klaus-Hendrik WOLF^a,
Reinhold HAUX^a, Uwe TEGTBUR^b

^a*Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig-
Institute of Technology and Hannover Medical School, Germany*

^b*Hannover Medical School, Institute for Sportsmedicine, Germany*

Abstract. The prevalence of overweight and obesity has increased worldwide in the last years as well as in Germany. The goal of cyberMarathon is to increase daily activities and participation of children and adolescents on sport courses using health-enabling technologies. Our objective is to propose an approach of an interdisciplinary intervention program containing a concept for architecture for a sensor-enhanced health information system which will be evaluate in two studies. Using sensors and health-enabling technologies in the preparatory study daily physical activities increased by 7,7%. Measuring of daily physical activities, the feedback of the analyzed data and the deal with the own body and activity seems to be an effective prevention for adolescents.

Keywords. Telemedicine, patient monitoring, systems architecture, compliance, pediatrics

Introduction

Current studies and physical examinations in schools show an overweight of 10-12% and an obesity of 4-6% of children [1], [2]. Obesity in the childhood causes problems like hypertension, cardiovascular diseases, diabetes etc [3]. The lack of physical activities and excessive inactivity (e.g. caused by watching TV) might enhance overweight and obesity in childhood [3]. The prevention and treatment involves dietary strategies and being more physically active. The most effective programs are family-based and/or school-based approaches to reduce obesity [4].

The goal of this paper is the presentation and discussion of a new approach called cyberMarathon which contains a concept for architecture for a sensor-enhanced health information system and an overweight prevention program developed by the Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig-Institute of Technology, in cooperation with the Institute for Sportsmedicine, Hannover Medical School, supported by the Federal Sports organization Lower Saxony.

¹ Corresponding Author: Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig-Institute of Technology, Germany, Muehlenpfordtstr. 23, 38106 Braunschweig, Germany; E-mail: maik.plischke@plri.de

Using new technologies is an approach to increase transparency concerning daily activities to improve the self-management of children. The target groups consists of children aged between 11 and 13 years because in this age the children leave sports clubs and decrease the daily physical activities like scientifically proven [5].

We considered the following research questions:

- Q1: How can health-enabling technologies increase the physical daily activities?
- Q2: How can health-enabling technologies be integrated in a prevention program?

1. Methods

Research Question Q1 is answered by describing the cyberMarathon approach. The intervention implemented within cyberMarathon is based on several prevention strategies against overweight and obesity in childhood and experiences in sensor-based telemedicine.

Research Question Q2 is answered by a description of the architecture for a sensor-enhanced health information system. We propose an architectural model using the three-layer graph-based metamodel (3LGM²)-methodology for static modeling of health information systems [6].

For the integration of sensors in prevention program with the possibility to feed back the analyzed data to children, parents, physicians and coaches we had to accomplish different challenges. Bott et al. [7] describe the problems of an integration gap concerning a personalized health information system (which describes the system in the personal environment at home or in school) and the different institutional health information systems (e.g. in hospitals, sports medicine examination centers,...). They conclude that it is possible to increase the patient empowerment and the self-management. Additionally there has to be a change in the personal health management and therefore a challenge in new ways of healthy living beside the problems of architectural and technical aspects. Also Mattila et al. [8] define several functions which have to be considered if health-enabling technologies will be implemented. The conclusion is that the trustful and volitional utilization of health-enabling technologies is one of the important challenges concerning the implementation of sensor-enhanced health information systems.

2. Results

2.1. The cyberMarathon approach

The goal of cyberMarathon is to increase the daily physical activities as well as the participation in sport courses of children and adolescents. A special attention applies thereby for pupils with an increased obesity risk.

New technology structures, called health-enabling technologies, are used and developed by computer scientists in cooperation with schools, sports clubs and physicians. The goals will be achieved by a new sports and movement program as well as by incentives to increase daily activities in combination with an improved health behavior. Latest sensor technology for measuring energy rates and activity is used in

connection with computer programs to improve the self-management of children. The sensor data which represent the daily activities have to be analyzed and stored in an electronic activity diary. Changes in energy rates and activities can be identified and the success could be visualized and presented to the children.

Additionally the sports program will be offered after school. On the basis of entrance tests physicians and coaches develop the sports and movement programs for the children. A group of one or more coaches trains the children for approx. 4-6 weeks once a week depending upon season e.g. inliner, dancing etc. Constant practicing could guarantee that unsporty children learn the kind of sport in such a way that they can control their success and decide, whether they want to continue the sports program further on. After those 4-6 weeks cycles the children take part in a one-week match. This can consist e.g. of the fact that the children are equipped over several days with a sensor for measuring daily activities. In the end the group with the highest activity level or the best metabolic rates wins.

The cyberMarathon approach contains of two scopes (Fig. 1):

1. The personal environment with the pupils at home or in school and different sensors to measure the activities and
2. The clinical environment with physicians, sports scientists etc.

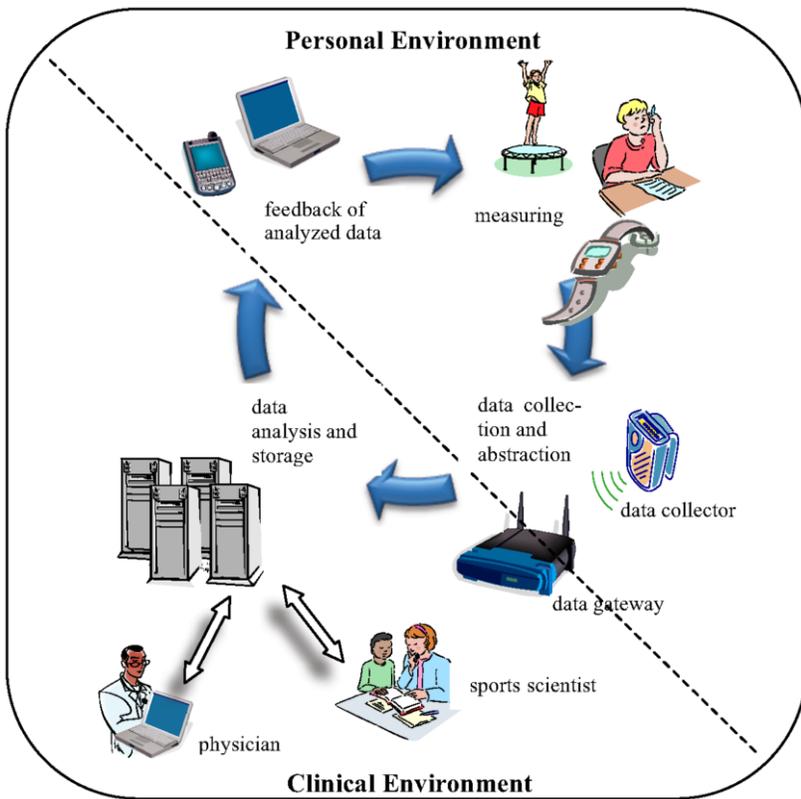


Figure 1. The cyberMarathon concept

2.2. A concept for the cyberMarathon architecture

Figure 2 shows the 3LGM² model of the architecture for the cyberMarathon approach described in the chapter before, with all entities, functions and relationships subdivided into three distinct architectural layers: the domain layer, the logical and the physical tool layer.

In the physical tool layer the different devices are illustrated described in the two scopes. In the personal environment the measuring and data collection takes place, in the clinical environment we have the possibilities to analyze the sensor data and the storage system for the activity diary.

All data are collected by the “data collector” and will be analyzed by an “ad hoc module” to allow a real-time feedback. The data are transferred by a gateway to the diary. The “diary” and the “analyzing module” in the logical tool layer are part of the electronic health record used by patients and physicians. The model architecture also contains a second loop of feedback to the patient: beyond the real-time feedback another loop with reports is available via an interface, e.g. internet browser.

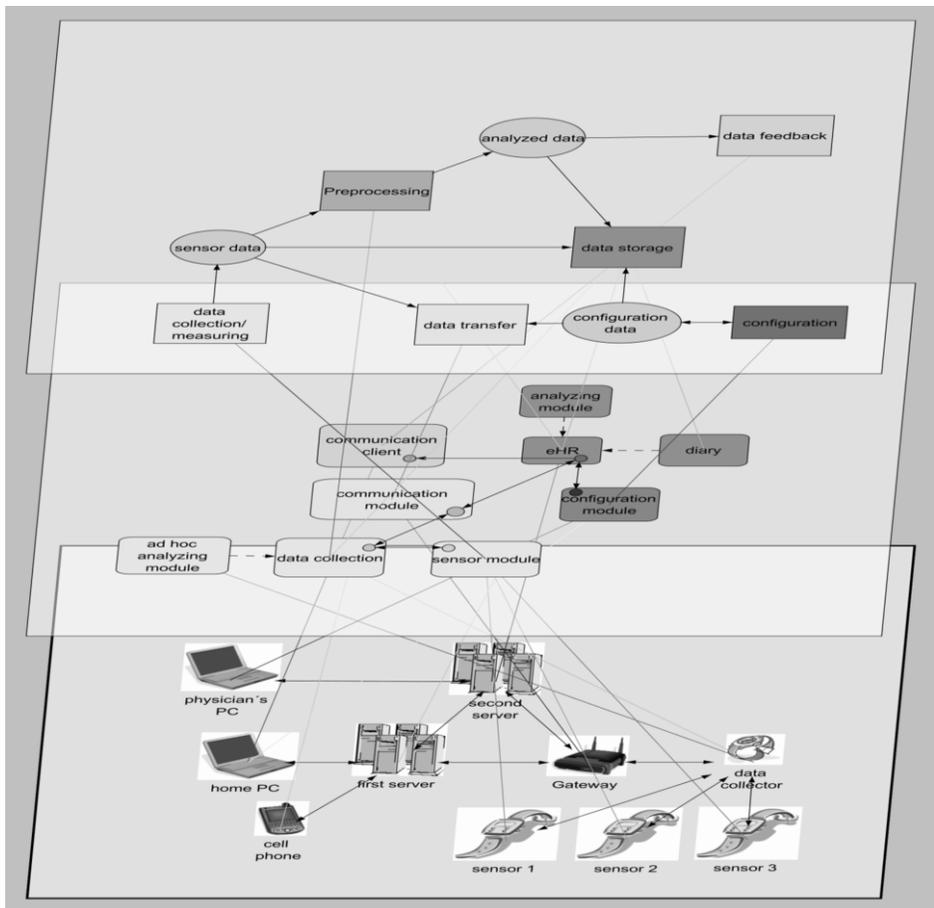


Figure 2: The three-layer graph-based metamodel (3LGM²) of the proposed cyberMarathon concept.

2.3. Case studies

With the preparatory study we have evaluated the measuring of data and the possibilities to feed back analyzed data described in the cyberMarathon architecture. In the preparatory study our focus of attention was the compliance of the children using this sensor and their problems understanding the presented measured and analyzed data. We decided to evaluate the effects of this intervention consisting additional sport courses and new technologies in a 6th grade school class.

We decided to use the sensor SenseWear Pro₂ distributed by Bodymedia to measure daily physical activities. The multi-sensor enables a personalized configuration by using weight, height, age, sex, left- or right-handed and smoker or non-smoker for calculation. The sensor measures acceleration, temperature of skin and air temperature close by the body, galvanic skin response and heat flow of the skin. The analyzing software calculates the parameter metabolic equivalent, activity duration, lying down, sleeping, number of steps and activity level. Over six months we had an increase of daily activities by 7,7%.

A second study with a sample of 44 children was realized from September 2006 to July 2007. The children were randomly sampled and examined in a school nearby Hannover, Germany. We chose the 6th grade of school forms which corresponds to children aged between 11 and 13 years. The main focus of attention of the second study was the investigation whether sensors and sensor-enhanced health information system have an effect on the health status measured by the body-mass-index. Over nine months we had an increase of daily activities by 11,4%. The results will be analyzed and published soon.

3. Discussion

Self-management is one of the key aspects to increase physical activities. Nevertheless the integration of the family is necessary, too. The correlation between the behavior of children and adolescents and the behavior of their parents is very noticeable [9]. The described support of self-management using health-enabling technologies [7], [10] could exemplary be implemented with cyberMarathon. Due to the fact that the inactivity of parents could affect the inactivity of their children it is important to implement a school-based intervention with coaches, teachers and e.g. scientists to motivate the children.

The cyberMarathon approach shows in a small setting that the compliance and acceptance of health-enabling technologies can improve healthy living. As well as other prevention strategies and studies like FITOC [11] or Jump-In [12] the cyberMarathon approach increases the physical activities. Instead of the often used method of increasing the number of sport lessons we have achieved this goal by increasing the transparency into the body and the daily activities using sensor technology. Bravata et al. achieved similar results using pedometers [13].

The task of prevention programs to clarify and educate children and parents as requested by the German workgroup for obesity in childhood and adolescence [14] can be supported by these technologies. Mattila et al. [8] define five categories for applications in pervasive computing. Among fitness and wellness cyberMarathon could be assigned to prevention and risk management. Subgroups of these categories are technical challenges described by Bott et al. [7] and challenges to prove efficiency and

effectiveness of health-enabling technologies [8]. Implementation of cyberMarathon and the results of the first study could contribute to prove the effectiveness.

4. Conclusion

There is considerable research in prevention and intervention of overweight and obesity in youth. Health-enabling technologies can support the different programs and studies concerning educational aspects. More transparency of the own body and more understanding about energy consumption, physical activities and the ratio of energy and movement increases patient empowerment and self-management of care and activity. The utilization of these new technologies seems to be accepted by the test persons over a long time period. This paper shows one possibility to integrate sensors and a sensor-enhanced health information system in an interdisciplinary intervention. Further research is needed for the optimal connection between sensor and the different infrastructures, the fusion of different data and the analysis algorithms.

References

- [1] Kromeyer-Hauschild K, Wabitsch M. Aktuelle Sicht der Prävalenz und Epidemiologie von Übergewicht und Adipositas bei Kindern und Jugendlichen in Deutschland: Leitlinien der Deutschen Adipositas Gesellschaft
- [2] Kurth BM, Schaffrath Rosario A. Die Verbreitung von Übergewicht und Adipositas bei Kindern und Jugendlichen in Deutschland - Ergebnisse des Kinder- und Jugendgesundheits surveys (KiGGS). Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz 2007; Springer, 50:736-743
- [3] Pinhas-Haniel O et al. Increased incidence of non-insulin-dependent diabetes mellitus among adolescents. J Pediatrics 1996; 128:608-15
- [4] Salmon J, Booth ML, Phongsavan P et al. Promoting Physical Activity Participation among children and adolescents. Epidemiol Rev 2007; 29:144-159
- [5] Lampert T, Mensink GBM, Romahn N, Woll A. Körperlich-sportliche Aktivität von Kindern und Jugendlichen in Deutschland - Ergebnisse des Kinder- und Jugendgesundheits surveys (KiGGS). Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz 2007; Springer, 50:634-642
- [6] Winter A, Brigl B, Wendt T. Modeling hospital information systems. Part 1: The revised three-layer graph-based meta model 3LGM2. Methods Inf Med 2003;42(5):544-51.
- [7] Bott OJ, Marscholke M, Wolf KH, Haux R: Towards New Scopes: Sensor Enhanced Regional Health Information Systems - Part 1: Architectural Challenges. Methods Inf Med. 2007; 46(4):476-483
- [8] Mattila E, Korhonen I, Saranummi N. Mobile and personal health and wellness management systems. In: Bradram JK, Mihailidis, Wan D, Pervasive Computing in Healthcare.CRC Press 2007; 105-134
- [9] Ebbeling CA, Pawlak DB, Ludwig DS. Childhood obesity: public-health crisis,common sense cure. Lancet 2002; 360:473-482
- [10] Haux R. Individualization, globalization and health—about sustainable information technologies and the aim of medical informatics. Int. J. Med. Inform 2006; 75:795-808.
- [11] Korsten-Reck U, Kromeyer-Hauschild K, Wolfarth B, Dickhuth HH, Berg A. Freiburg Intervention Trial for Obese Children (FITOC): results of a clinical observation study. International Journal of Obesity 2005; 29:356-361
- [12] Jurg MW, Kremers SPJ et al. A controlled trial of a school-based environmental intervention to improve physical activity in Dutch children: JUMP-in, kids in motion. Health Promotion International 2006; 21(4): 320-330
- [13] Bravata DM et al. Using Pedometers to increase physical activity and improve health. JAMA 298. 2007; 2296-2304
- [14] Leitlinien der Arbeitsgemeinschaft Adipositas im Kindes-und Jugendalter (AGA) www.a-g-a.de, last access: 14.11.2007

Integrating clinical, gene expression, protein expression and preanalytical data for *in silico* cancer research

Delphine ROSSILLE ^{a,c,1}, Anita BURGUN ^{a,c}, Céline PANGAULT-LORHO ^b
and Thierry FEST ^b

^a*UPRES EA3888 Conceptual Modeling of Biomedical Knowledge,
Rennes University, France*

^b*INSERM U917 Microenvironnement & Cancer, HITC Department,
Rennes Hospital, France*

^c*Medical Information Department, Rennes Hospital, France*

Abstract : We present the phase I development of an integrative platform for the analysis of clinical, gene expression, protein expression and pre-analytical data. The platform is aimed at providing transparent access and analysis tools to researchers investigating new biomarkers and prognosis factors in the particular field of lymphoma diseases. In this article, we report on the data integration phase. The platform's principal advantage is its completeness as it integrates in a single environment clinical, genomic and proteomic data, allowing for their combined analysis. The architecture consists in a data warehouse including data on patients, clinical trials and array platforms and a DeMilitarized Zone for data exchange. A secure web-based platform allows any collaborative team to request the data warehouse and access basic statistics on integrated data. The presented system is currently in use.

Keywords : Systems integration; Information storage and retrieval; Databases, Genetic; Cancer

Introduction

Nowadays, an increasing amount of genomic and annotation data is produced, stored worldwide and analyzed for better understanding of diseases and investigating new biomarkers. After independently studying genomic, proteomic or clinical data it has become clear that simultaneous analysis, by increasing the analysis power, is the next step. However biologists are faced with the problem of exploiting such massive heterogeneous and distributed data. They need systems to transparently access relevant data and to facilitate their interpretation thanks to appropriate analysis tools.

The BMS-Ly national research project, conducted by Rennes hospital, is devoted to investigating blood biomarkers and prognosis factors of the Diffuse Large B-Cell

¹ Corresponding Author : Delphine ROSSILLE, Département d'Information Médicale, CHU Rennes, 2 rue Henri Le Guilloux, 35033 Rennes, France, Email : delphine.rossille@chu-rennes.fr

non-Hodgkin Lymphoma (DLBCL) disease, more specifically predictors of answers to treatments. The study is based on data collected at the diagnosis, that is before patients received any treatment. Our contribution to this project aims at supporting biostatisticians by setting up an integrative platform for the analysis of all the produced data (clinical, gene and protein expression data and preanalytical parameters). We report here on the phase I development involving the data integration. Phase II will be concerned with the development of tailored analysis methods, based on statistics and data mining.

1. Background

Bioinformatics requires the development of informatics systems to facilitate the handling, querying and exploration of high throughput data. The largest project is probably CaCore [1] that provides data standards and system infrastructures for cancer research. Since the explosion of high-throughput technologies, many tailored data management systems have been developed. Some are deposit and retrieval systems for published well-annotated experimental data like [2] for gene expression data of any species and pathology or [3] for gene and protein expression data. Others are specific to a pathology ([4] for clinical and gene expression data of solid cancers) or a medical domain ([5] for gene and protein expression data for immune cells). Few integrate gene and protein expression data [3, 6]. Some public data repository and retrieval systems provide in addition online analysis tools (Array Express with the Expression Profiler online tool [7], ITTACA [8] with bioinformatics analysis for clinical and gene expression data). Only few laboratory databases include clinical, gene and protein expression data in their systems for complex analysis ([9], [10]).

2. Material and Method

2.1. Material

The cohort will consist in 300 patients and 150 healthy control subjects. The patients' dataset is collected from the GOELAMS (Groupe Ouest-Est des Leucémies aigües et Autres Maladies du Sang) 02-03 and 075 clinical trials, two French nationwide multicenter studies launched in 2005. The local medical doctor asks the GOELAMS for the inclusion of a new patient. Having checked the inclusion criteria, the GOELAMS trial manager provides an anonymized inclusion number to be used from then on. The local medical doctor accesses the electronic Case Report Form (eCRF) online system including built-in password security. The collected data include clinical, demographic, biological data, received treatments, abnormal events or randomization information and are stored in a database in Nancy. The blood samples collected at the inclusion stage are processed in Rennes before being sent for transcriptomic or proteomic analysis.

Preanalytical data, including the delays between the sampling and freezing processes, are collected by Rennes to check their impact on the samples quality [11]. The transcriptomic data are produced by Montpellier Affymetrix microarray platform and sent to Rennes in their raw proprietary format (.cell files). They are derived from Affymetrix Human Exon 1.0 ST arrays Hu-Ex-1_0-st, arrays providing the most

comprehensive coverage of the genome, including not only well annotated regions. The protein expression data are produced by Grenoble Bio-Rad SELDI-TOF Mass Spectrometry platform. Produced spectra, detected peaks and experimental conditions are sent to Rennes. All experimental data are related to their patients thanks to an anonymization number.

2.2. Method

As the patients' data are distributed in different data sources, an appropriate system's architecture was determined in accordance with required functionalities. The database conceptual modeling was developed so as not to be restricted to clinical trials, the investigators wishing for future use to store data from patients having the same pathology but not included in the clinical trial. The platform was then implemented.

3. Results

3.1. Required Functionalities

The study involves five French research teams that will actively collaborate. Four teams produce the clinical, gene and protein expression and pre-analytical data. Two teams will investigate the data. The data will be analyzed with commercial software (on raw expression data files) or with tailored analysis methods (based on statistical and data mining methods), hence data must be available in a unique format. Finally the Rennes hospital requested all data to be banked for future research.

3.2. System's Architecture

We chose to integrate all the data into a tailored data warehouse, for centralizing into a unique system with a unified data schema. This approach, compared to federated or mediation systems, ensures consistency among data, rapid queries on massive amount of data and fast complex analyses. The main drawback is when data updates are required. However in our case, only exceptional updates could be needed and all transactions are historically recorded in a log book. Data elements are integrated in the data warehouse and files of raw experimental data and experimental conditions are stored in a data server and linked to the anonymized inclusion numbers of their patients. The data warehouse includes array platforms and clinical trial information.

The platform is a secure web site accessible by all the teams (Figure 1). A DMZ zone is used to download by secure FTPs the produced data files, for their latter integration into the database. Gene annotations are accessible by hyperlinks to public databases from the probeset identifier for transcriptomic data.

Platform: PhP, HTML , javascript - Data warehouse: Oracle 9i relational database.

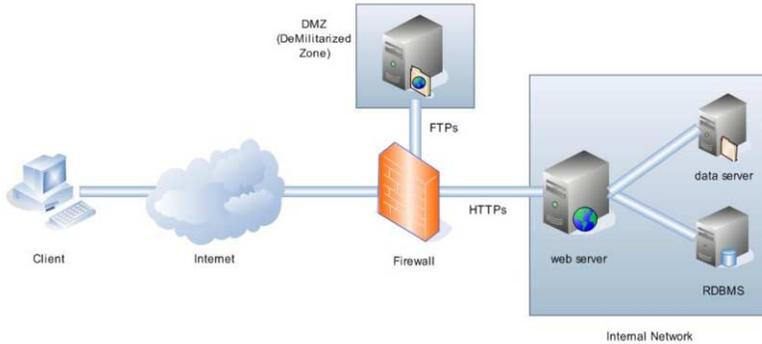


Figure1: System's architecture

3.3. Conceptual Data Modeling

The model (Figure 2) consists in six data spaces: four patient-related (clinical, gene expression, protein expression and pre-analytical data spaces), the clinical trials and the array platforms spaces. In the future other spaces (publications, gene annotations or biomarker results) could easily become part of the model. The modeling is generic enough to integrate patients not included in clinical trials.

Hence a patient is integrated for a medical *episode* related to a specific *disease*, an *episode* consists in several *stages*. At each stage the patient can have *treatments*, *exams results*, pre-analytical and expression data or *events* such as status aggravation. For instance, in the case of the GOELAMS trial, $\langle typeEpisode \rangle$ is 'clinical trial' and the first $\langle typeStage \rangle$ is 'inclusion' to which are associated pre-analytical parameters, gene and protein expression data and the clinical data collected at the inclusion stage. Pre-analytical data are associated to expression data as it informs on the sample quality (Table *PreAnalyticalParameters*). Each *TranscData* is related to a probeset identifier (attribute *id_probe*), and an *ArrayPlatform* consists in a list of probeset identifiers as labelled by the manufacturer (table *probesets*), and general information on its design name (attribute *nameArray*), manufacturer, version and species. The raw files of each genomic or proteomic analysis are linked to its experimental conditions (for instance attribute *transcRawData_fileName* in table *TranscExperimentalConditions*).

3.4. The Platform's functionalities

- All data, except for proteomic data (not yet implemented), are automatically integrated into the database from formatted files. Clinical data must be integrated first in order to create the patients. The gene expression data can only be integrated if the corresponding array platform is registered into the database.
- Descriptive statistics of the database status are displayed according to criteria s.a. sex or age, allowing the analyser to check data integration over time.
- Data for patients' groups can be exported according to clinical criteria. Hence combined analysis of expression and clinical data is made possible. The corresponding list of raw data files is produced and allows their analysis through commercial software.
- Authentication and access rights : In addition to data anonymization, the secure web site and the DMZ are accessible through Rennes hospital's firewall with login and password authentication (Figure 1). Access rights to the web site are either for "user"

or “administrator”, the last one allowing to integrate data into the database, register new array platforms, reschedule integration of past files and check the log book.

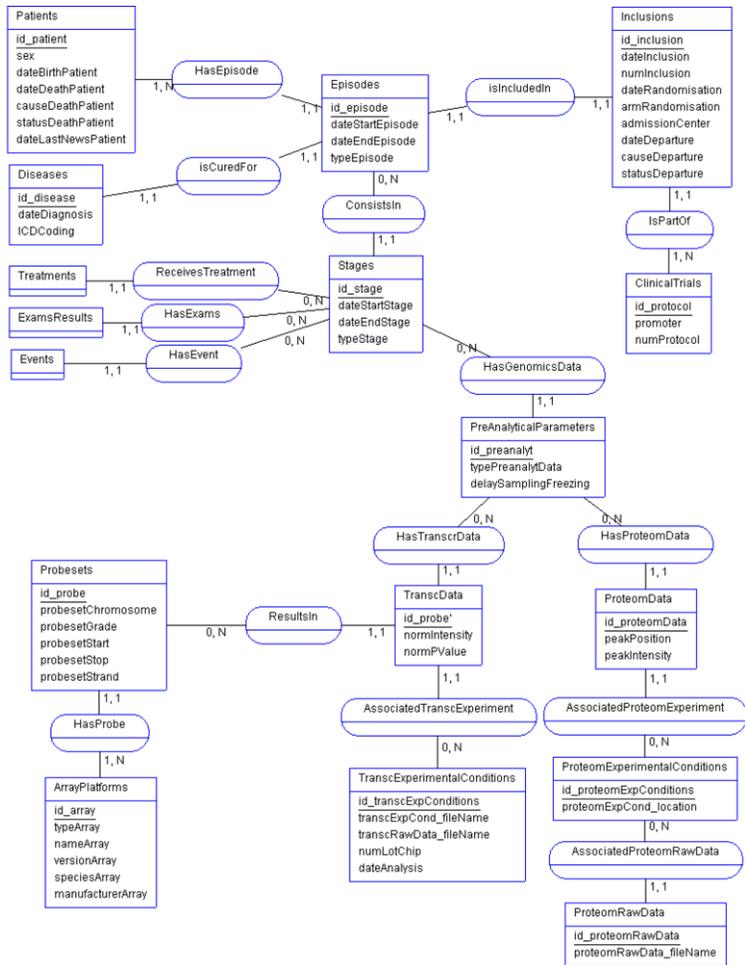


Figure 2: Data Modeling

4. Discussion and Conclusion

The national French BMS-Ly research project was funded to discover biomarkers from blood samples of patients diagnosed with DLBCL. We present the data integration platform built to support researchers by making data access transparent regardless of their heterogeneity. As of today, the presented platform is in use, the integration of protein expression data remaining to be implemented. Security of patients’ data has been taken care of: anonymized inclusion numbers, firewall and secure architecture. Phase II will focus on the development of tailored analysis tools.

In regards of the literature on similar platforms [2-8], by integrating pre-analytical, clinical, genomic and proteomic data in a unified data model, the platform's main advantage is its data completeness. Indeed the majority of published platforms store gene or protein expression data, to the exception of [3,5-6]. Furthermore clinical data are usually limited in scope. Our application integrates all available clinical data from the eCRF system. It includes information on the samples quality that can be taken into account during analyses. It integrates full exon expression data not all well-annotated or even not annotated at all, providing finer insights. The presented data warehouse is aimed not to be limited to its current application but will be most valuable in the future for simultaneous analysis of gene and protein expression data. Furthermore, contrary to public applications like [8] where genomic data are stored in files and only a description is kept in the database, the presented platform stores data as files and as data elements, making the focus on specific clinical and expression data elements simple. GeWare [12] is a similar integrative platform, with the difference of the data model used (multidimensional vs relational).

The presented platform can be improved. For instance, data on experimental conditions will be in the future compliant to standards, such as the MIAME standard for transcriptomic data. The platform has several limitations. Presently, the data to be integrated are well formatted: the clinical data are mostly based on controlled vocabularies specified by the GOELAMS, expression data are stored in proprietary formats. The syntax variations detected for some clinical data were taken care of. Semantic interoperability will however have to be overcome when patient's data will be derived from other sources. Ontologies like SNOMED CT, by offering the ability to relate concepts together, will be then most valuable. Finally the platform should accommodate cross-platform integration and analysis.

References

- [1] P.A. Covitz *et al*, caCORE: A common infrastructure for cancer informatics, *Bioinformatics* **19** (18) (2003), 2404-2412
- [2] H. Parkinson *et al*, ArrayExpress – a public repository for microarray gene expression data at the EBI, *Nucleic Acids Research* **33** (2005), D553-D555
- [3] K. Ikeo *et al*, CIBEX: Center for Information Biology gene Expression database, *C.R.Biologies* **326** (2003), 1079-1082
- [4] K. Kato *et al*, Cancer gene expression database (CGED): a database for gene expression profiling with accompanying clinical information of human cancer tissues, *Nucleic Acids Research* **33** (2005), D533-D536
- [5] A. Hijikata *et al*, Construction of an open-access database that integrates cross-reference information from the transcriptome and proteome of immune cells, *Bioinformatics* **Sept. 25** (2007)
- [6] C. Chelala *et al*, Pancreatic Expression database: a generic model for the organization, integration and mining of complex cancer datasets. *BMC Genomics* 2007, 8:439
- [7] M; Kapushesky *et al*, Expression Profiler: next generation – an online platform for analysis of microarray data, *Nucleic Acids Research* **32** (2004), W465-W470
- [8] A. Elfilali *et al*, ITTACA: a new database for integrated tumor transcriptome array and clinical data analysis, *Nucleic Acids Research* **34** (2006), D613-D616
- [9] H. Hu *et al*, Biomedical informatics: development of a comprehensive data warehouse for clinical and genomic breast cancer research, *Pharmacogenomics* **5** (7) (2004), 933-941
- [10] A. Nagarajan *et al*, Database challenges in the integration of biomedical data sets, In *Proc. 30th VLDB Conference*, Toronto, Canada, 2004, 1202-1213
- [11] C. Pangault *et al*, Stakes of pre-analytical parameters in blood transcriptomic and proteomic analysis: Application to clinical research: The GOELAMS trial, *Med Sci (Paris)* **3** (2007), Mar 23, 13-18(french)
- [12] E. Rahm, The GeWare data warehouse platform for the analysis of molecular-biological and clinical data, *J. Integrative Bioinformatics*, **4** (1) (2007), journal.imbio.de

Developing a taxonomy of communication errors in heterogeneous information systems

Samrend SABOOR^{a,1}, Elske AMMENWERTH^a

^a *Institute for Health Information Systems, UMIT - University for Health Sciences, Medical Informatics and Technology, Hall in Tyrol, Austria*

Abstract: Background: Although established communication standards do exist in the health care domain (i.e., DICOM and HL7) the communication within heterogeneous information systems still shows a variety of errors. The complexity of these systems aggravates the identification of error reasons. A structured summary of communication errors and their reasons is essential for developing methods that support the error detection. Methods: In order to summarize communication errors, a systematic literature review in PubMed was conducted. Selected references were filtered iteratively and analyzed by applying subsuming qualitative content analysis. Results: The taxonomy currently contains 12 different problem classes that group 42 problems, with in total 130 reasons. Discussion: Although, not all selected literature references are yet analyzed, we observe a saturation concerning new errors/error classes. In order to increase validity and completeness expert interviews are in planning stage. However, the first results are promising.

Keywords: Classification, HIS management, Systems architecture

1. Introduction

The correct transmission of information objects between involved computer-based application systems (e.g., order entry system) has become vital for processes in health care institutions [1]. Involved communication partners must agree upon conventions in order to cooperate effectively [2-4]. These conventions pertain the structure of information objects (e.g., as sets of identifier-attributes pairs), the meaning of each attribute and consequently each application system's communication interfaces.

Although there are established communication standards in the health care domain (i.e., DICOM (Digital Imaging and Communications in Medicine, [5]) and HL7 (Health Level 7, [6])) the communication within heterogeneous information systems is still error-prone [7-9]. Even the usage of just one of the standards requires additional implementation efforts [10, 11]. A main reason for this is because standard definitions allow misinterpretations. Additionally, in case of communications including different standards, difficulties in matching the particular information models can lead to

¹ Corresponding Author: Samrend Saboor. University for Health Sciences, Medical Informatics and Technology (UMIT) – Eduard Wallnöfer-Zentrum 1, 6060 Hall in Tyrol, Austria. Samrend.Saboor@umit.at

fundamental content-related errors. Therefore, an additional framework is needed that coordinates the usage of the different communication standards [12, 13]. This framework is provided by the international initiative IHE (Integrating the Healthcare Enterprise).

However, integrated hospital information systems mostly are grown architectures that include proprietary workarounds. Changes in these systems mostly succeed because of experienced IT-employees. On the other side, communications which appear successful can still have mistakes in the exchanged content. In these cases, the underlying reasons are mostly difficult to identify. Further, even changes in well working infrastructures may cause negative side-effects which are hardly predictable (e.g. [8, 13]). Problems related to communication infrastructure and processes affect the quality of patient treatment and must be examined carefully – detecting communication errors in advance.

However, it seems that there is no present method that aims to assess health care communication processes. Existing methods mainly focus on time measurements (e.g., MOSAIK-M [14]) or perform reachability analyses on Petri-net based models in order to detect bottlenecks and best performing variations [15]. These methods either ignore information objects or provide only simple representation per default. In the latter case, information objects are not defined on a formal base and thus cannot be utilized by automated assessment methods. Thus, a method is needed that considers the characteristics of information objects and their processing in heterogeneous information systems. Here, an important prerequisite is a concise summary of weaknesses which can occur within the communication between computer-based application systems.

Aim of this paper: This paper presents a taxonomy of weaknesses that can occur within the communication between computer-based application systems.

2. Methods

Based on our experiences from earlier projects in the area of process assessment (e.g., [16]), we conducted a systematic review of available literature in PubMed in order to collect communication errors and their underlying reasons. Figure 1 illustrates the inductive approach we chose: After the initial declaration of the review's aim, we derived search phrases from this aim in the second step – starting with simple search phrases like “quality information processing” or “experience information processing”. In steps three and four the title and abstract of all references which resulted from the search phrases were reviewed. In step five, we adjusted and augmented our search phrases according to the adequateness of the resulting references. In this way, we found 4188 references. These references dealt amongst others with information management (e.g., [17]), reports on integration projects in the field of Hospital Information System (HIS), Radiology Information System (RIS) and Picture Archiving and Communications Systems (PACS) (e.g., [18, 19]). From this set we dropped all references which were older than 20 years, dealt with the implementation of very specialized software applications or dealt with organizational issues. In the sixth step, we performed qualitative content analyses on the remaining 426 references. Here, we chose a subsuming type of qualitative content analysis (according to Mayring [20]) which aims to filter the main content by abstraction and dynamic declaration of categories. In the sixth step, also further keywords were found which we used in step

seven to adjust our set of search phrases. The process stops when saturation of new errors is reached.

The found weaknesses were collected in a taxonomy which is presented in Table 1 in the Results section.

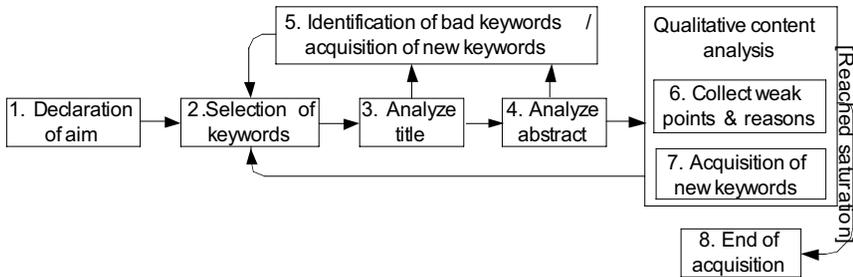


Figure 1: Process of literature review – After declaring the review’s aim, appropriate keywords and references are searched iteratively. Adequate results are analyzed for communication weak points and their reasons. The process stops after an observable saturation

3. Results

The found communication errors were collected in a taxonomy – a three-level-hierarchy: It groups communication errors into separate problem classes and names reasons for the occurrence of each error. Table 1 shows an excerpt of the taxonomy. Currently, the taxonomy contains 12 problem classes. In these, 42 problems are grouped. For these problems 130, partly redundant, reasons were collected. Additionally, 49 problems with 62 reasons are already collected but not yet inserted into the taxonomy. However, a saturation for the problem classes and the actual problems is already observable. Further details regarding the taxonomy’s content as well as concepts of how to make use of it will be presented on the MIE conference 2008.

Table 1: Excerpt of the taxonomy of communication errors

Problem class	Problem	Reason for problem
Input errors	Wrong data entries	Mistakes due to manual data entry, missing entry checking routines, no usage of automated data gathering routines
Identification errors	Missing or wrong identification of information object instances	Incompatible identification attributes between requester and provider, wrong or missing assignment of identification, several systems inconsistently assign identifications to same information object
Conversation errors	Mistakable conversation	Not all details are transferred via electronic communication paths
Incomplete content	Values of important attributes are missing	Optional / textual data fields are missing or filled with wrong values, important details are filled into optional data fields, missing identification of essential data fields, important details are passed verbally
	Acquisition errors	See problem class "Acquisition errors"
	Input errors	See problem class "Input errors"

Conversion errors	Incompatible value representation of data attributes	Usage of data attributes with implicit value representation, conversion error in communication standard, different length encodings, different data dictionaries, incompatible character sets
	Incompatible data models	Usage of proprietary information objects / dependency on proprietary attributes, usage of proprietary communication protocols, different/incompatible information models, usage of obsolete communication standard versions,
	Missing content	See problem "Values of important attributes are missing"
Consistency errors	Uncoordinated data entry	See problem class "Input errors", multiple independent input-interfaces, multiple independent/not synchronized data bases
	Missing data cleansing	No central data repository, missing central data correction
	Corrupted main repository	No quality assurance procedures for main data repository
	Delayed failure correction	Defective information objects are re-sent with delays
	Redundant data management	No main repository or missing connections to established repository, one-way connections just allow data synchronization in one direction, redundant data entries in one or more repositories
	No version management	Storage of incomplete versions of single information object instances
Transmission errors	Unsupported services	The called application system doesn't support the requested service, usage of outdated application systems, new version of application system has other communication interfaces
	Unsupported content	The called application system doesn't support the request type of information object, receiving system stores transferred information object differently than the original (different information model)
	Missing processing rules	Missing trigger for continuation of information processing
	Unstable software versions	Usage of unstable/experimental application systems
	Restrictive security setup	Firewall blocks legal communications
	Malfunctions of network	No alternative network-components (missing uninterrupted power supply unit, switch, cable), no security redundancies
	Communication disruption / Incomplete communication	Communication partners are not connected directly (no synchronous comparison), peer misses to reply instantly
	Communication with wrong partner	Ambiguous identification of application systems
	Conversion errors	See problem class "Conversion errors"
Availability errors	Missing authorization	No or wrongly configured authorization prohibits access, information objects are not stored on central servers - no distributed access
	Acquisition/storage errors	See problem class "Incomplete content", shortage of storage space
	Missing tracking of information objects	Multiple duplications, storages and transmissions of information object instances, paper-based information objects easily disappear
	Incompatible/Missing or wrong identification	See problem class "Identification errors"
	Insufficient network bandwidth	No direct connection - data must be transferred offline via storage media, usage of mixed storage media - necessity of transcriptions
	Concurring access	Usage of paper-based storage media, paper-based information objects not available for distributed access
Dataloss/	Physical data corruption	System crash or improper system shut-down, sensitive storage media

damage errors	Missing backup	Missing redundant data storage/secure long-term-archive, no usage of RAID-technology, no dedicated backup-server
Acquisition errors	Error during acquisition	Missing final visual control/quality assurance, disruptions during acquisition, exporting of incomplete information objects, no structure data entry forms

4. Discussion

Communication processes in health information systems are complex interactions between multiple application systems with different and possibly diverse interpretation of common communication standards. Errors in the electronic communication of clinical documents can negatively effect patient treatment and thus must be detected in advance. Here, a structured summary of such errors and their reasons is the essential prerequisite for the development of adequate detection methods – these methods could then be used, for instance, by the hospital’s information management department during the planning stage of new integration projects but also befor updating central application systems.

The taxonomy presented in this paper is meant to be such a collection. Its details were gathered through a qualitative content analysis on available literature that deals with experience reports about integration projects and the implementation of common communication standards (i.e., HL7 and DICOM). Although, not all selected literature references are yet analyzed, we observe a saturation concerning new errors/error classes. Thus, the taxonomy seems to be nearly finished. However, the subsuming qualitative approach we selected for gathering the taxonomy’s details strongly depends on the selected literature references. This selection in turn inherently contains a certain degree of subjectivity resulting from subjectively selected search phrases. Amongst others, this affects the validity and completeness of our taxonomy. Thus, we tried to reduce this subjectivity by iteratively refining the search phrases according to the found literature resources. Further, we are currently planning expert interviews in order to increase the taxonomy’s validity and completeness. However, the already collected results are encouraging. As far as we can see, a concise summarization of concrete communication such as the taxonomy is still missing – current publications (e.g., [17]) mostly concentrate on single problems or discuss them on an abstract level.

5. Conclusion

It is difficult to understand the reasons for communication errors and to overlook possible side-effects within integrated information systems. A structured taxonomy of possible errors and their reasons is an essential prerequisite for solving this difficulty. In this paper we presented such a taxonomy which groups communication errors into problem classes and also names reasons for each of the errors. In order to parameterize this taxonomy, we currently develop a formal method. The method’s concept is explained in [21] in more detail.

6. References

1. Coiera E. When conversation is better than computation. *JAMIA* 2000;7(3):277-286.
2. Brazhnik O, Jones JF. Anatomy of data integration. *J Biomed Inform* 2007;40(3):252-69.

3. Giuse DA, Kuhn KA. Health information systems challenges: the Heidelberg conference and the future. *Int J Med Inform* 2003;69(2-3):105-14.
4. Kun L. Interoperability: the cure for what ails us. *IEEE Eng Med Biol Mag* 2007;26(1):87-90.
5. ACR/NEMA. DICOM Homepage. 2007 [cited 2007 Jun]; Available from: <http://medical.nema.org>
6. HL7. Health Level 7. 2007 [cited 2007 Jun]; Available from: <http://hl7.org>
7. Hammond KW, Helbig ST, Benson CC, Brathwaite-Sketoe BM. Are electronic medical records trustworthy? Observations on copying, pasting and duplication. *AMIA Annu Symp Proc* 2003:269-73.
8. Nissen-Meyer S, Holzknacht N, Wieser B, Francke M, Küttner B, Adelhard K, et al. Produktivitätsverbesserung durch klinikweite RIS und PACS: eine Fallstudie. *Der Radiologe* 2002;42:351-60.
9. Khorasani R. Medical management: expanding radiologists' role using information technology to improve the quality of care. *Semin Roentgenol* 2003;38(3):282-6.
10. Gross-Fengels W, Miedeck C, Siemens P, Appel R, Muckner K, Finsterbusch J, et al. [PACS: from project to reality. Report of experiences on full digitalisation of the radiology department of a major hospital]. *Radiologe* 2002;42(2):119-24.
11. Matthews JW, Bosch WR. Explicit-VR transfer syntax limits the value multiplicity of DICOM data elements with decimal string (DS) value representation. *Phys Med Biol* 2006;51(5):L11-2.
12. HIMMS/RSNA. IHE - changing the way healthcare connects. 2007 [cited 2007 Jun]; Available from: <http://www.ihe.net>
13. Mildenerger P, Wein B, Bursig HP, Eichelberg M. [Current developments in DICOM and IHE]. *Radiologe* 2005;45(8):682-9.
14. Hoffmann I, Bergmann J, Bott OJ, Pretschner DP. Einsatz einer rechnergestützten Modellierungs- und Simulationsumgebung für den Entwurf telemedizinischer Systeme am Beispiel von MOSAIK-M. In: Steyer G, Tolxdorff T, editors. *TELEMED*; 2005 2005; Berlin: Aka-Verlag; 2005. p. 309-20.
15. Blake J, Carter M, O'Brien-Pallas L, McGillis-Hall L. A formal model for quality assurance in coloproctologic surgery. In: Greenes R, Peterson H, Protti D, editors. *MEDINFO 1995, Proceedings of the 8th World Congress on Medical Informatics*. Amsterdam: North Holland; 1995. p. 527-31.
16. Saboor S, Chimiak-Opoka J, Ammenwerth E. Supporting the systematic assessment of clinical processes: the MedFlow method. *Methods Inf Med* 2007;46(5):586-594.
17. Haux R, Winter A, Ammenwerth E, Brigl B. *Strategic Information Management in Hospitals*. New York, USA: Springer-Verlag; 2004.
18. Mongkolwat P, Bhalodia P, Gehl JA, Channin DS. Validating DICOM content in a remote storage model. *J Digit Imaging* 2005;18(1):37-41.
19. Kotter E, Langer M. Integrating HIS-RIS-PACS: the Freiburg experience. *Eur Radiol* 1998;8(9):1707-18.
20. Mayring P. *Einführung in die qualitative Sozialforschung: eine Anleitung zu qualitativem Denken*. 5., überarb. Aufl. ed. Weinheim: Beltz, Psychologie-Verl.-Union; 2002.
21. Saboor S, Ammenwerth E. Assessing communication processes within integrated health information systems. Accepted for publication in: *Proceedings of the Sixth IASTED International Conference on Biomedical Engineering*; 2008; Innsbruck, Austria; 2008.

Mining Knowledge from Corpora: an Application to Retrieval and Indexing

Lina F.SOUALMIA^{a,1}, Badisse DAHAMNA^b, Stéfan DARMONI^b

^a *LIM&Bio EA 3969, SMBH Léonard de Vinci, Paris XIII University, Bobigny, France.*

^b *CISMeF team and LITIS EA 4051, University of Rouen, France*

Abstract. The present work aims at discovering new associations between medical concepts to be exploited as input in retrieval and indexing. *Material and Methods:* Association rules method is applied to documents. The process is carried out on three major document categories referring to e-health information consumers: health professionals, students and lay people. Association rules evaluation is founded on statistical measures combined with domain knowledge. *Results:* Association rules represent existing relations between medical concepts (60.62%) and new knowledge (54.21%). Based on observations, 463 expert rules are defined by medical librarians for retrieval and indexing. *Conclusions:* Association rules bear out existing relations, produce new knowledge and support users and indexers in document retrieval and indexing.

Keywords. Data Analysis; Indexing; Terminology; Data Mining; MeSH.

Introduction

Internet is a major source of biomedical knowledge. As the access to structured medical information is difficult with directories or general search engines, many applications have been developed [1]. Since 1995, CISMeF (acronym of Catalog and Index of French-speaking Medical Sites) [2] has been selecting institutional and educational resources for patients, students and health professionals. It references 36,247 e-documents by using Medical Subject Headings (MeSH) [3]. Among many sources to support users, such as morphological bases, dynamic and contextual search tools [4], MeSH structure is exploited. To complete these sources, we propose to mine e-documents to discover new associations between medical concepts by data mining.

1. Material and Methods

1.1. Medical Subject Headings, Resource Types and Metaterms

The MeSH thesaurus is used by the National Library of Medicine for indexing biomedical resources. Its core is a hierarchical structure that consists of sets of

¹ Corresponding Author: Lina Soualmia, LIM&Bio EA 3969, 74 rue Marcel Cachin, 93017 Bobigny, France; E-mail: lina.soualmia@gmail.com.

descriptors: at the top level general headings (e.g. *diseases*) and deeper more specific headings (e.g. *brain infraction*). The 2007 version contains over 24,357 main headings (e.g.: *hepatitis*) and 83 subheadings (e.g.: *diagnosis*). Together with a main heading, a subheading can be used to specify a particular aspect. For example, the pair [*hepatitis/diagnosis*] specifies *diagnosis* aspect of *hepatitis*.

MeSH is originally used to index biomedical scientific articles for the MEDLINE database. In order to customize it to the field of e-health resources resource types have been introduced [2]. CISMef resource types are an extension of MEDLINE publication types (e.g. *clinical guidelines*). Each document in CISMef is described with a set of MeSH main headings, subheadings and CISMef resource types. Each main heading, [main heading/subheading] pair and resource type is allotted a ‘minor’ or ‘major’ weight, according to the importance of the concept it refers to in the resource. Major terms are marked by a star (*).

1.2. Data Mining

Knowledge extraction from databases or data mining in computer science [5] consists in discovering additional information from large structured sets of data. This knowledge could be used to do predictions about new data or to explain existing data. One of the objectives of extraction process is the generation of association rules. It is processed in several steps: data and context preparation (objects and items selection), extraction of *frequent itemsets* (compared to a minimum support threshold), generation of *most informative rules* using a data mining algorithm, and finally interpretation and deduction of new knowledge [6]. An extraction context is a triplet $C=(O, I, R)$ where: O is the set of objects, I is the set of all the items and R is a binary relation between O and I .

1.2.1. Association Rules

A data mining system may generate several thousands and even several millions frequent association rules, and only some of them are interesting. An association rule is interesting if it is easily understandable by the users, valid for new data, useful or if it confirms a hypothesis. It is expressed as: $i_1 \wedge i_2 \wedge \dots \wedge i_k \Rightarrow i_{k+1} \wedge \dots \wedge i_n$ and states that if an object has the items $\{i_1, i_2, \dots, i_k\}$ it tends also to have the items $\{i_{k+1}, \dots, i_n\}$. *Support* represents the rule utility. It corresponds to the proportion of objects which contains at the same time antecedent and consequent. $Support = |\{i_1, i_2, \dots, i_n\}|$. *Confidence* represents precision and corresponds to the proportion of objects that contains the consequent rule among those containing the antecedent. Two rule types are distinguished: exact rule having Confidence=100%, i.e. verified in all the objects of the database and approximative rule. $Confidence = |\{i_1, i_2, \dots, i_n\}| / |\{i_1, i_2, \dots, i_k\}|$.

1.2.2. A-Close for Mining e-Documents

The problem of the relevance and the usefulness of extracted association rules is of a primary importance because real-life databases lead to several thousands and even millions of association rules whose confidence measures are high, and among which are many redundancies, i.e. rules conveying the same information among them. Two bases for association rules are defined by A-Close [7]. These bases generate sets for all valid non-redundant association rules, being thus smaller, composed by minimal

antecedents and maximal consequents i.e. the most relevant association rules. We adapt A-Close to the case of e-health documents data base by considering conceptual indexing: the set of objects O is the set of indexed documents; the set of items I is the set of MeSH descriptors; the relation R represents the indexing relation between an object and an item, i.e. between a document and a descriptor.

1.2.3. Processing Collections of Documents

End-users are categorised in CISMef in mainly three types: professionals, students in medicine, patients and lay people. Rather than extracting knowledge referring to the main medical specialties as in [4], we consider the three major resource types *guidelines**, *education** and *patients** and two kinds of itemsets: the set of major main headings (MH*) and the set of major [main heading/subheading] pairs (MH/SH*).

Table 1. Description of the collections of documents.

Resources	Documents	Items	Min	Max	Mean
Guidelines*	2,727	MH*	1	64	5.21
		MH/SH*	1	70	6.12
Patients*	3,272	MH*	0	25	1.63
		MH/SH*	0	30	1.82
Education*	3,610	MH*	0	25	2.22
		MH/SH*	0	34	2.73

2. Results

2.1. Mining e-Documents

For all contexts, minimum support was fixed to $\text{minsup}=20$ and minimum confidence to $\text{minconf}=70\%$ for the approximative association rules (Table 2). We obtain association rules between major MH* (resp. MH/SH* pairs). For the major resource types *patients** and *education** all (100%) association rules are between two MHs* (resp. MH/SH* pairs) i.e one descriptor in the antecedent and one descriptor in the consequent. For *guidelines** 24% of the rules are between more than two descriptors. Characteristics of documents may explain these results: average descriptors from 1.63 to 2.22 for *patients** and *education** whereas 5.21 to 6.12 for *guidelines**.

Table 2. Number of rules, exact rules (ER), approximative rules (AR) and pairs.

Resources	Context: item=MH*				Context: item=[MH/SH]*			
	Rules	ER Conf=1	AR Conf \geq 0.7	Pairs	Rules	ER Conf=1	AR Conf \geq 0.7	Pairs
Guidelines*	50	12 (24%)	38 (76%)	38 (76%)	39	8 (20.51%)	31 (79.49%)	35 (76%)
Patients*	20	9 (45%)	11 (55%)	20 (100%)	19	8 (42.1%)	11 (57.9%)	19 (100%)
Education*	23	6 (26.09%)	17 (73.91%)	23 (100%)	25	13 (52%)	12 (48%)	25 (100%)

Another experiment is carried out in the context of documents with the resource type *guidelines** to obtain more complete association rules: we consider the descriptors MH and MH/SH pairs without allotted minor or major weight. An average of 12

descriptors with a minimum of 1 and a maximum of 301 descriptors compose the documents (Table 3). As A-Close works on databases with a maximum of 12 items, we have added a constraint on the number of descriptors. To avoid long time generation and to have interpretable association rules, we added the maximum size of the closed itemsets as a new parameter of the algorithm.

Table 3. Description of the documents of the Guidelines* collection.

	Docs	Items	Min	Max	Mean
Guidelines*	2,727	MH	1	111	10.08
		MH/SH	1	301	13.54

We obtain a high number of association rules with a minimum support threshold $minsup=20$ and a minimum confidence threshold $minconf=70\%$ (Table 4) but only 0.95% (respectively 1.92%) are between two MH (respectively between two MH/SH pairs). By reducing the confidence from 1 to 0.7 the number of rules between MH (respectively between MH/SH) grows with a factor of 5 (respectively 4.42).

Table 4. Association rules between MH and MH/SH in the context Guidelines*.

Items	Rules	ER Conf=1	AR Conf≥0.7	Pairs
MH	35,454	6,990 (19.71%)	28,464 (80.29%)	338 (0.95%)
MH/SH	27,011	6,102 (22.6%)	20,909 (77.4%)	520 (1.92%)

2.2. Association Rules Evaluation

As defined, an interesting association rule confirms a hypothesis or states a new hypothesis [6]. We propose here to combine background domain knowledge with simple statistical measures used traditionally in association rules mining for evaluation. We consider several cases of interesting association rules according to relations between MeSH descriptors [3]. As these relations are defined between two main headings and between two subheadings we consider only the association rules between two elements. Hence, an interesting existing association rule could associate: a (in)direct son and its father (FS); two descriptors that belong to the same hierarchy (same (in)direct father) (B); two descriptors with See Also relation (SA). These rules are automatically classified thanks to MeSH structure. The other rules that satisfy the $misup$ and $minconf$ are then considered as «new» interesting association rules.

Exact association rules, except for collection *patients**, are mostly new interesting rules: from 62.5% to 99.86%. Therefore, existing rules are mainly from the *patients** collection: 77.77% for MH* and 75% for MH/SH*. Approximative rules, except for the *guidelines** collection with items MH and MH/SH pairs, are mostly existing interesting rules: from 58.07% to 78.73%. New interesting rules are between MH and MH/SH from the collection *guidelines**: 99.73% for MH and 99.52% for MH/SH.

Subjective interest measures are based on the expert knowledge about the data, i.e. here the medical librarian. New interesting rules for the contexts MH* and MH/SH*

Figure 1. Some examples of new interesting rules validated by the expert

breast cancer/diagnosis → mammography
aids/prevention and control → condom
influenza vaccines → influenza/prevention and control
Turner syndrome ∧ child → human growth hormone ∧ growth disorders
obstetric delivery → pregnancy
prostate cancer/surgery → biopsy ∧ prostatectomy
amniocentesis → prenatal diagnosis ∧ chorionic villi sampling
opioids analgesics/administration and dosage → pain/drug therapy

pairs are evaluated manually. 93.75% (resp. 84.78%) of the interesting new rules with confidence=1 (resp. confidence≥0.7) between major descriptors are validated.

Table 5. Association rules evaluation according to MeSH structure.

		Exact rules: Confidence=1				Approximative rules: Confidence≥0.7			
		Existing knowledge			New	Existing knowledge			New
		FS	B	SA		FS	B	SA	
Patients*	MH*	0 0%	5 55.55%	2 22.22%	2 22.22%	2 18.18%	2 18.18%	4 36.36%	3 27.27%
	MH/SH*	0 0%	5 62.5%	1 12.5%	2 25%	2 18.18%	2 18.18%	3 27.27%	7 36.36%
Education*	MH*	1 16.66%	1 16.66%	0 0%	4 66.67%	2 11.76%	6 35.29%	3 17.64%	6 35.29%
	MH/SH*	1 7.69%	0 0%	1 7.69%	11 87.62%	2 16.76%	3 25%	2 16.76%	5 41.66%
Guidelines*	MH*	0 0%	0 0%	4 33.33%	8 66.67%	2 5.26%	7 18.42%	10 26.31%	12 31.57%
	MH/SH*	1 12.5%	1 12.5%	1 12.5%	5 62.5%	3 9.67%	3 9.67%	9 29.03%	13 41.93%
	MH	0 0%	2 0.03%	8 0.14%	6,980 99.86%	12 0.04%	37 0.13%	30 0.10%	28,382 99.73%
	MH/SH	6 0.1%	4 0.06%	7 0.11%	6,085 99.73%	25 0.12%	50 0.23%	27 0.13%	20,807 99.52%

2.3. Indexing Correction and Expert Rules

Documents are manually indexed and according to the indexing policy, the more precise descriptor should be used, i.e. in lower level in hierarchy. However, 1,466 documents contain descriptors that have father-son relation and 478 documents are indexed by subheadings that have a relation while associated to the same keyword. For example, a document is indexed by *trisomy* and *chromosome aberrations*, whereas *trisomy* is a *chromosome aberration*. This may explain the proportion of existing associations. Correction should be proposed to the indexers.

Main return on experiences of association rules extraction and evaluation is modeling and formalisation of rules between [main heading/subheading] pairs based on observations. The pattern of the rule *hepatitis/prevention and control*→*hepatitis vaccines* is used to model *dysentery bacillary/prevention and control* → *shigella vaccines*. 463 rules are modeled. Formalization concerns different cases and contexts for retrieval and indexing. The rule $MH_1/SH_1 \xrightarrow{-} MH_2$ states that MH_1/SH_1 should be replaced by the main heading MH_2 . For example *abdomen/radiography*→*radiography abdominal*. The rule $MH_1/SH_1 \xrightarrow{++} MH_2/SH_2$ states that the pair MH_2/SH_2 should be added to the pair MH_1/SH_1 . *appendectomy*→*appendicitis/surgery* states that the pair *appendicitis/surgery*

should be added to queries (or to document description when indexing) containing the main heading *appendectomy*.

3. Discussion and Future Work

There is an increasing activity in text mining in the genomic model [8]. In [9] co-occurrences between Gene Ontology terms are analyzed and association rules are mined to identify pairs of related Go Terms. Association rules are more complete than co-occurrences measures between pairs of concepts but one of the challenging issues is the overabundance of associations that may be discovered as in [10]. A-Close generates all the valid non-redundant association rules composed by minimal antecedents and maximal consequents. Evaluation is processed in two steps: first the selection of the most informative rules and second the classification of the rules according to the MeSH taxonomy structure to filter existing associations. Only the most frequent rules that are not classified are presented to the expert for a final evaluation. This method combines statistical measures and background domain knowledge.

Association rules are used in retrieval by query expansion (automatic and interactive) and enriching users' queries with new knowledge [4]. As exact rules (respectively approximative rules) state that the antecedent and the consequent are at the same time in all (respectively some) documents, this kind of rules should be used in automatic (respectively interactive) query expansion. However, these expansions work only in the case of queries that return documents. Association rules link conceptual structures of the documents i.e. descriptors organised in hierarchies on which it is possible to make specialization and generalization. We plan to generate generalized association rules and to examine how other data collections such as MEDLINE will work with our approach. Association rules and expert rules can be translated in the form of automatons for processing automatic indexing of raw text documents. Finally formalised association rules could improve the power of reasoning based on MeSH-OWL [11].

References

- [1] Abad Garcia F et al. A comparative study of six European databases of medically oriented Web resources. *J Med Libr Assoc.* 2005;93(4):467-479.
- [2] Douyère M et al. Enhancing the MeSH thesaurus to retrieve French online health resources in a quality-controlled gateway. *Health Info Libr J.* 2004;21(4):253-261.
- [3] Nelson SJ et al. Relationships in MeSH. *Kluwer Publishers* 2001;171-84.
- [4] Soualmia LF, Darmoni SJ. Combining knowledge-based methods to refine and expand queries in medicine. *LN in Computer Science* 2004;3055:243-255.
- [5] Agrawal R, Srikant R. Fast algorithms for mining association rules in large databases. *Very Large Data Bases* 1994;478-499.
- [6] Fayyad UM et al. Advances in knowledge discovery and data mining. Am. Ass. Artificial Intelligence Press 1996;601-611.
- [7] Lakhal L et al. Efficient mining of association rules using closed itemset lattices. *Information Systems* 1999;24:25-46.
- [8] Ananiadou S, Mc Naught J. Text mining for biology and biomedicine. *Artech House publishers* 2005.
- [9] Bodenreider O et al. Non lexical approaches to identifying associative relations in the Gene ontology. *Pacific Symposium on Biocomputing* 2005;10:91:102
- [10] Berardi M et al. A data mining approach to PubMed query refinement. DEXA 2004. *IEEE Computer Society*; 401-405.
- [11] Soualmia LF et al. Representing the MeSH in OWL: towards a semi-automatic migration. *KR-Med* 2004;72-80.

Enhanced Information Retrieval from Narrative German-language Clinical Text Documents using Automated Document Classification

Stephan SPAT ^{a,1}, Bruno CADONNA ^a, Ivo RAKOVAC ^a, Christian GÜTL ^b,
Hubert LEITNER ^c, Günther STARK ^c and Peter BECK ^a

^a *Institute of Medical Technologies and Health Management,
JOANNEUM RESEARCH Forschungsgesellschaft mbH, Graz, Austria*

^b *Institute for Information Systems and Computer Media,
Graz University of Technology, Graz, Austria*

^c *Steiermärkische Krankenanstaltenges. m.b.H., Graz, Austria*

Abstract. The amount of narrative clinical text documents stored in Electronic Patient Records (EPR) of Hospital Information Systems is increasing. Physicians spend a lot of time finding relevant patient-related information for medical decision making in these clinical text documents. Thus, efficient and topical retrieval of relevant patient-related information is an important task in an EPR system. This paper describes the prototype of a medical information retrieval system (MIRS) for clinical text documents. The open-source information retrieval framework Apache Lucene has been used to implement the prototype of the MIRS. Additionally, a multi-label classification system based on the open-source data mining framework WEKA generates metadata from the clinical text document set. The metadata is used for influencing the rank order of documents retrieved by physicians. Combining information retrieval and automated document classification offers an enhanced approach to let physicians and in the near future patients define their information needs for information stored in an EPR. The system has been designed as a J2EE Web-application. First findings are based on a sample of 18,000 unstructured, clinical text documents written in German.

Keywords. EPR-CPR-EMR, Classification, Data analysis-extraction tools

Introduction

The amount of digital textual information in hospitals is increasing. Especially the implementation of Electronic Patient Record (EPR) systems, which

¹ Corresponding Author: Stephan Spat, JOANNEUM RESEARCH Forschungsgesellschaft mbH, Elisabethstrasse 11a, 8010 Graz, AUSTRIA, stephan.spat@joanneum.at

manage the patient related electronic information, have reinforced the demand for finding relevant information, particularly in unstructured clinical text documents.

Steiermärkische Krankenanstalten Ges.m.b.H. (KAGes), the governing body of Styrian hospitals, covers 20 hospitals with about 6,000 beds and 16,000 employees for over 1.2 million inhabitants. In order to displace heterogeneous IT-systems of numerous hospitals by offering an integrative Hospital Information System (HIS), in 2004, the roll-out of a new HIS – termed openMEDOCS – was conducted. openMEDOCS is the synonym for the implementation of the software packages IS-H from SAP and i.s.h.med from GSD and T-Systems. The core of openMEDOCS is an EPR system which stores all documents of patients generated in KAGes hospitals. [1-2]

Considerable amounts of managed data in the EPR system are narrative clinical text documents. Since the roll-out of openMedocs, the amount of clinical text documents has increased continuously. Since no effective retrieval tool is available yet, physicians spend a lot of time finding relevant information in patient records. Thus, efficient and effective retrieval of relevant patient-related information stored in these documents has gained importance. Consequently, a project leader of openMEDOCS asked for a prototype of a medical information retrieval system (MIRS).

Clinical text documents like discharge letters or reports are highly relevant for physicians' decision-making processes regarding patient-care. These text documents are often narrative, non-structured, with few metadata. Approaches for standardizing narrative texts have been developed with HL7 CDA [3], however, they are mostly not implemented in clinical practice yet and do not necessarily include metadata needed for efficient information retrieval. Thus, text-mining may be a viable method for finding relevant information in non-structured documents [4-8].

In this paper, we describe an enhanced approach combining metadata generation by means of *automated document classification* with standard *information retrieval* techniques, which we implemented in a prototypical medical information retrieval system (MIRS).

1. Materials and Methods

In order to provide physicians with a tool to improve the search capabilities in the EPR system of openMEDOCS regarding efficiency and effectiveness of finding relevant patient-related information, this chapter describes the design of the prototypical MIRS.

Figure 1 illustrates the overall design: narrative clinical documents are stored in the modeled EPR system ('EPRs'). For indexing, searching, and classification, these documents are gathered from EPRs by the 'DB Component'.

The 'Classification Component' trains and evaluates classification models for classifying unseen narrative clinical text documents into an arbitrary number of predefined categories (i.e. medical fields).

The 'Index Component' extracts index terms from the documents and stores them together with additional metadata like document name, date of last modification as well as medical fields predicted by the 'Classification Component' in the 'Index'.

Physicians define their information needs using the 'User Interface (UI) Component'. The user question (search string and metadata) is analyzed and terms are extracted. Furthermore, the medical field categories generated by document classification can be used influencing the rank order of retrieved documents. In the

following, a query is written including this information. The query is used to search within the 'Index'.

Finally, a ranked document list with snippets of the document content is returned to the user. If the snippet of the document presented to the user has aroused interest, the original document may be opened.

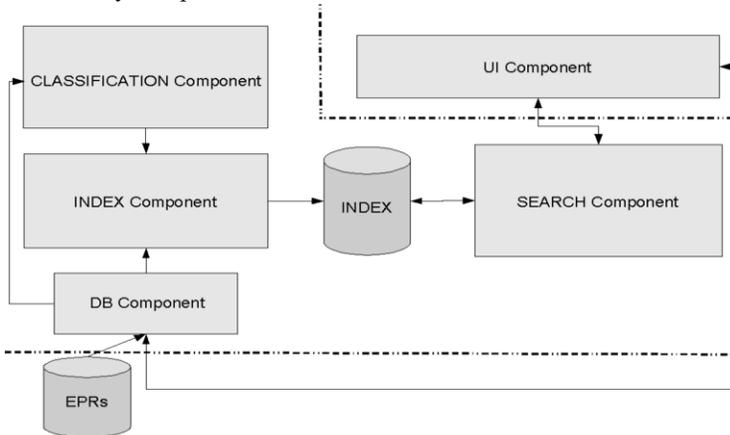


Figure 1. Overall design of the MIRS.

1.1. EPRs/DB Component

A simplified EPR system was modeled for the prototype to store and process text documents extracted from openMEDOCS' EPR system. 18,000 narrative clinical text documents written in German were extracted. The extracted set comprises 26 different types of clinical documents (e.g. discharge letters, reports) from eight medical fields (surgery, vascular surgery, casualty surgery, internal medicine, neurology, anaesthesia and intensive care, radiology and physiotherapy). Documents were provided as plain text. Each document can be associated with a specific patient.

1.2. Classification Component

For the purpose of classifying documents in medical fields, a *multi-label text classification system* [9] using the open-source data mining framework WEKA [10] was developed. The classification system was trained with a sample of 1,500 text documents manually classified by a domain expert (physician).

Four different classification models were compared for the classification task: J48 – a tree based classifier, SMO – an implementation of support vector machines, k-NN – an instance based classifier, and Naïve Bayes – a classifier based on probabilistic theory [11-14]. For each model, a 10-fold cross validation was conducted. The F1-measure [15] was used as comparable evaluation measure.

Multi-label classification was used because information in a document can refer to multiple medical fields. For example, a person who had had an accident and had broken her leg was first in the radiology to X-ray her leg. Afterwards, she had an operation in the medical department of casualty surgery. All stages of treatment were documented in a single document generated by the department of casualty surgery but

the information in the document refers to the medical fields ‘radiology’ and ‘casualty surgery’. Multi-label classification allows classifying a document in multiple medical field categories.

1.3. Index and Search Components

The open-source information retrieval framework Apache Lucene [16] was used for the implementation of the indexing and searching components of the medical information retrieval system (MIRS) prototype. Apache Lucene allows setting a ‘boost factor’ which influences the relevance score of retrieved information. If the user retrieves information with particular attention to specific medical fields, documents classified in these medical fields will have a high relevance score pushing the documents to the top of the results list.

1.4. J2EE Web Application

The system was designed as a J2EE web application which has the following advantages: (1) Platform Independence, (2) Multi-Access, (3) Modularity/Re-usability, and (4) Data Security. For future applications, the design allows patients to access their own medical records with an ordinary web-browser.

2. Results and Discussion

2.1. Automated Document Classification

Four different classification schemes have been compared for the classification task: J48, SMO, k-NN, and Naïve Bayes. Since evaluation showed J48 classification model to achieve best results with a F1 measure of 0.89 [9], it has been used to classify clinical documents in the developed MIRS prototype.

2.2. Sample Application of MIRS

In the following, a sample application of the MIRS prototype and especially the effect of ‘relevance boosting’ will be illustrated.

A physician is interested in particulars about medical treatments of the ‘herz’ (heart) of a patient with ID 12019922. The user enters the user query ‘herz*’ into the search mask. The use of the wildcard ‘*’ expands this term to all words which start with ‘herz’. Additionally, the physician is particularly interested in documents which belong to the medical fields ‘Chirurgie’ (surgery) and ‘Innere Medizin’ (internal medicine). Therefore, the user sets the medical field booster to ‘Chirurgie’ and ‘Innere Medizin’.

After the submission of the query settings, the MIRS prototype first looks for all documents which contain at least one word which starts with ‘herz’. In a second step, ‘relevance boosting’ applies, i.e. the MIRS calculates relevance scores for found documents resulting in higher scores for documents which were classified as ‘Chirurgie’ or ‘Innere Medizin’ and lower scores for those which were not classified in those classes. The corresponding list of results shows documents ordered by relevance

score in descending order. It is important to note that documents which were not classified into one of the two classes are not excluded from the list of results because the classification might have been erroneous or incomplete. They can be found at the bottom of the list of results.

Figure 2 shows an extract of the list of results. The document at the first position in the list was classified as ‘Chirurgie’ and ‘Innere Medizin’, which results in the highest relevance score. The following documents were assigned just to one of the chosen classes resulting in lower relevance scores. On the bottom of the list (not shown) are documents which are not assigned to any of the classes chosen by the physician.

	Document preview	Last modification	Document type	Predicted medical fields	Score
show	Herzschrittmacher in situ Coronarer Bypass+ Klappe 2003....Mitralklappenersatz. Das Herz ist gut tonisiert, allseits grenzwertig groß beidseits E78.5 Hyperlipidämie onÄ links Z95.0 Herzschrittmacher in situ I25.1 Koronare Herzkrankheit Gefäß onÄ I10 Hypertonie onÄ links Z95.0 Herzschrittmacher in situ Ulcus an der Cardia	28/06/2005	Ärztlicher Bericht Med	[Chirurgie, Innere Medizin]	5
show	Herzschrittmacher in situ I25.1 Koronare Herzkrankheit Gefäß onÄ I10 Hypertonie onÄ links Z95.0 Herzschrittmacher in situ Ulcus an der Cardia	16/08/2005	Anaesthesie Präoperativer Check	[Chirurgie, Anaesthesie/Intensiv]	4
show	Herzschrittmacher in situ I25.1 Koronare Herzkrankheit Gefäß onÄ I10 Hypertonie onÄ links Z95.0 Herzschrittmacher in situ Ulcus an der Cardia	10/08/2006	Anaesthesie Präoperativer Check	[Chirurgie, Anaesthesie/Intensiv]	3
show	Herzschrittmacher in situ I25.1 Koronare Herzkrankheit Gefäß onÄ I10 Hypertonie onÄ links Z95.0 Herzschrittmacher in situ Ulcus an der Cardia	12/09/2005	Verlegungsbericht Anaesthesie	[Chirurgie, Anaesthesie/Intensiv]	2
show	Herzschrittmacher in situ I25.1 Koronare Herzkrankheit Gefäß onÄ I10 Hypertonie onÄ links Z95.0 Herzschrittmacher in situ Ulcus an der Cardia	26/02/2005	Anaesthesie Präoperativer Check	[Chirurgie, Anaesthesie/Intensiv]	1

Figure 2. Snippet of the resulting document list of ‘relevance-boosted’ documents.

In a first evaluation, four searching tasks – conducting two tasks with and two tasks without ‘relevance boosting’ – were given to five clinicians who are using openMEDOCS in their practical routine. Qualitative feedback emphasizes that the MIRS could improve search speed and search quality in clinical text documents. The clinicians also underlined that ‘relevance boosting’ speed up searching, especially in patient records with many text documents.

The implemented prototype shows an approach to influence the rank order of retrieved document using automated multi-label document classification. Therefore, documents have been classified into medical fields with good results.

As the implementation of MIRS is prototypical, functionality is restricted to basic search and index functions. Neither structured data, like diagnosis codes, nor other media like X-ray radiographs were considered. Moreover, profound evaluation of the prototype regarding usability, retrieval quality, and response time was left for future work.

3. Conclusion

The increasing number of narrative clinical text documents demands for natural language processing techniques like information retrieval or automated document classification [4]. Based on well established open source frameworks, this paper presents an enhanced approach combining both in one medical information retrieval system.

References

- [1] Kraßnitzer M. EDV im Spital: Der Patient auf Knopfdruck, *CliniCum*, 7-8/2006, Online; <http://www.medical-tribune.at/dynasite.cfm?dssid=4171&dsmid=74897&dspaid=582979>, last accessed: 2007-10-18.
- [2] Leitner H. openMEDOCS erfolgreich eingeführt, *G'sund.net*, 50/2006, Online; <http://www.gsund.net/cms/beitrag/10073293/2052790>, last accessed: 2007-11-15.
- [3] Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, Shabo Shvo A. HL7 Clinical Document Architecture, Release 2. *J Am Med Inform Assoc*. 2006 Jan–Feb; 13(1): 30–39.
- [4] Müller M, Markó K, Daumke P, Paetzold J, Roesner A, Klar R. Biomedical data mining in clinical routine: expanding the impact of hospital information systems. *Stud Health Technol Inform*. 2007;129:340-4.
- [5] Marko K, Schulz S, and Hahn U. MorphoSaurus—design and evaluation of an interlingua-based, cross-language document retrieval engine for the medical domain. *Methods Inf Med* 2005; 44(4): 537-545.
- [6] Ehrler F, Ruch P, Geissbuhler A, Lovis C. Challenges and methodology for indexing the computerized patient record. *Stud Health Technol Inform*. 2007;129:417-21.
- [7] Cimino JJ, Li J. Sharing infobuttons to resolve clinicians' information needs. *AMIA Annu Symp Proc*. 2003;:815.
- [8] Currie LM, Graham M, Allen M, Bakken S, Patel V, Cimino JJ. Clinical information needs in context: an observational study of clinicians while using a clinical information system. *AMIA Annu Symp Proc*. 2003;:190-4.
- [9] Spat S, Cadonna B, Rakovac I, Gütl C, Leitner H, Stark G, Beck P. Multi-label text classification of German language medical documents. *Stud Health Technol Inform*. 2007;129:1460-1.
- [10] Witten, I. H., and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, second ed., Morgan Kaufmann, San Francisco, 2005.
- [11] Quinlan, JR. *C4.5: Programs for Machine Learning*, Morgan Kaufmann, 1993.
- [12] Platt J. Fast training of support vector machines using sequential minimal optimization, In B. Scholkopf B, Burges C and Smola A, *Advances in Kernel Methods - Support Vector Learning*, MIT Press, 1998.
- [13] Aha DW, Kibler D, Albert, MK. Instance-based learning algorithms, *Machine Learning*; 6(1): 37-66, 1991.
- [14] John G, Langley P. Estimating continuous distributions in bayesian classifiers, paper presented to Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, Vancouver, Canada, 1995.
- [15] Hripcsak G, Rothschild AS: Agreement, the F-Measure, and Reliability in Information Retrieval. *J Am Med Inform Assoc*. 2005; 12(3): 296-298.
- [16] Apache Lucene: <http://lucene.apache.org/>, last accessed: 2007-10-17.

Frequency of hospital-acquired pneumonia in electronic and paper-based patient record

Jürgen STAUSBERG ^{a,1}, Abdelouahid AZAOUAGH ^b

^a *Medical Faculty, University of Duisburg-Essen, Germany*

^b *Medizinische Klinik I, Sankt Marien-Hospital Mülheim a.d. Ruhr, Germany*

Abstract. Aim of the study was to check the validity of the electronic patient record for hospital-acquired pneumonia and to estimate its reliability. Reviewing 23,356 inpatients with admission from the 1st April 2005 and discharge up to the 30th September 2005 we identified 211 cases with hospital-acquired pneumonia in the electronic patient record whereof 70 cases taken at random were included in the validation. A second random sample of 130 cases under risk was used to calculate its completeness. For hospital-acquired pneumonia, the latest version of the definition of the Centers for Disease Control and Prevention (CDC) was applied. In 64.3 % of the cases hospital-acquired pneumonia was confirmed in the paper-based patient record (45 cases, 95%-confidence interval 51.9 % - 75.4 %). Beside 10 cases with pneumonia already existing with admission 15 cases remain where even pneumonia could not be confirmed in the paper-based patient record. Completeness was calculated as 42.9 %. Estimation for the University Clinics Essen revealed a true rate of hospital-acquired pneumonia of 1.32 % (308 of 23,356). The estimated true rate was higher than the measured rate of 0.90 % (211 of 23,356) which is covered, nevertheless, from the 95%-confidence interval. Data from the electronic patient record seem to be sufficient to forecast the true rate of hospital-acquired pneumonia, for example, for questions of infectious disease epidemiology. However, it is not sufficient enough for special claims of the supervision, clinical hygiene and prevention for which an optimization of the data quality is required.

Keywords. Centers for disease control and prevention, computerized patient records, hospital, incidence, nosocomial infections, paper-based patient record, pneumonia, routine data

Introduction

Pneumonia is one of the most frequent hospital-acquired infections and an important issue of patient safety [1]. It occurs at least 48 hours after admission to a hospital or 14 days after discharge. Main risk factors of hospital-acquired pneumonia are tracheal intubation, restricted consciousness, chronic lung disease, abdominal or thoracic operation, aspiration of fluid, and an age of more than 70 years [2]. The most important risk factor is a mechanical ventilation of more than 1 day at an intensive care unit.

¹ Corresponding Author: Jürgen Stausberg, MD, PhD, Kordulastr. 13, D-45131 Essen, Germany; E-mail: stausberg@ekmed.de.

Hospital-acquired pneumonias carry harm for patients and an economical burden for the health care system. A continuously surveillance is a main mean in its prevention. Prevention is considered as the most effective procedure to reduce the mortality related to hospital-acquired pneumonia. About one third of hospital-acquired pneumonias are regarded as preventable [3].

The use of routine data available in an electronic patient record (EPR) has got great attention in quality control and surveillance [4]. It could be shown that the secondary use of routine data is one possibility to improve quality of care through the introduction of information and communication technologies in health care [5]. Regular feedback can support a weakness analysis [6]; a timely notion of adverse events can trigger protocol guided actions [7]. Nevertheless, the quality of data recorded electronically in parallel to a paper-based patient record is discussed controversially [8]. One can assume that the data are tailored to the tasks supported by each type of record, e. g. legal issues with the paper-based type and reimbursement with the electronic type. In case of functional overlap it might be the user who decides about the particular type of record used for the storage of a single item.

In the University Clinics of Essen patient safety is an issue of high priority, e. g. in the field of pressure ulcer [9]. Focusing on hospital-acquired pneumonia the question rose, whether the quality of the data available in the EPR is sufficient enough to estimate the frequency of hospital-acquired pneumonia. In the following we will introduce the study design and its results.

Material and Methods

We included inpatients admitted from 1st April 2005 and discharged not later than 30th September 2005. The definition for hospital-acquired pneumonia in the electronically available routine data was as follows: a code for pneumonia from the German adaptation of the ICD-10 (ICD-10-GM 2005) **AND NOT** a code for pneumonia as admission and/or principal diagnosis. The period prevalence rate was then calculated as number of inpatients with hospital-acquired pneumonia as numerator and the total number of included inpatients as denominator. For the paper-based patient record we used the definition of the Centers for Disease Control and Prevention (CDC) [10]. The estimation of the true period prevalence rate is based on a combination of a) the period prevalence rate from the EPR adjusted for length of stay, b) the correctness of the EPR adjusted for length of stay and c) the completeness of the EPR. For the completeness we calculated the 95 %-confidence interval. The paper-based patient record was used as gold standard. Correctness and completeness were calculated as proposed by Hogan and Wagner [11].

For the calculation of correctness we took a random sample of inpatients with a hospital-acquired pneumonia in the EPR. Because we wanted to detect a difference of 20 % for correctness with an a priori figure of 80 %, sample size calculation leads to 70 inpatients (α -error 0.05, β -error 0.1). A priori we expected a completeness of 50 % and a true rate of hospital-acquired pneumonia of 20 % in a population at risk with a length of stay from at least 15 days and discharge from the Department of Anesthesiology, Department for Bone Marrow Transplantation, and the Center of Internal Medicine. The sample size calculation leads to 130 inpatients (α -error 0.05, β -error 0.1). Due to 6 inpatients present in both samples, 194 paper-based patient records remain for retrospective review concerning the presence of a hospital-acquired pneumonia

according to the CDC-criteria. The analysis of the paper-based patient records was done without notion of the EPR-status. All records could be retrieved.

The EPR at the University Clinics of Essen is used in parallel to the paper-based one. It includes for example lab results, radiology reports, and clinical data needed for reimbursement with the German diagnosis related groups. For the latter, diagnoses and procedures are coded with legislatively obliged classifications by physicians and/or documentation specialists. The paper-based patient record is still the legislative obliged archive of all relevant data. Due to missing mobile computers it is used at the point of care as well. Therefore, electronically generated data like lab results are printed out and stored in the paper-based record.

Results

From the sample of 70 inpatients, 45 hospital-acquired pneumonias were confirmed in the paper-based patient record. The correctness is 64.3 %. This is significantly lower than the expected rate of 80 % ($p=0.002$). 10 inpatients presented pneumonia already at the time of admission, in 15 inpatients pneumonia could not be confirmed at all in the paper-based patient record.

Within the high risk sample of 130 inpatients we identified 28 with hospital-acquired pneumonia. From these, 12 were already identified in the EPR. The completeness is 42.9 % (95 %-confidence interval 24.5 % - 62.8 %). Thus, completeness confirmed the expected range.

The estimation of the true period prevalence rate is shown in table 1. First, we calculated the correctness for each group of length of stay. Then, we estimated the number of inpatients for each group by multiplying the number of inpatients identified in the EPR with the correctness figure (i.e. for group 1: 41 inpatients \times 0.083 = 3.4 real cases). The resulting number of 132 inpatients (rounded) is combined with a completeness of 42.9 % (95 %-confidence interval 24.5 % - 62.8 %) leading to 308 (210 - 539) true inpatients suffering from a hospital-acquired pneumonia, i.e. a true period prevalence rate of 1.32 % (0.90 % - 2.31 %).

Table 1. Calculation of the true period prevalence rate.

	Short stay (1-5 days)	Normal stay A (6-10 days)	Normal stay B (11-20 days)	Long stay (> 20 days)
Inpatients	14,614	4,765	2,609	1,368
Inpatients with hospital-acquired pneumonia EPR	41	29	51	90
EPR period prevalence rate	0.3 %	0.6 %	2.0 %	6.6 %
Random sample	12	9	21	28
Confirmed in the paper-based patient record	1	3	16	25
Correctness	8.3 %	33.3 %	76.2 %	89.3 %
Estimated true number of inpatients with hospital-acquired pneumonia	3.4	9.7	38.9	80.4

Discussion

In the case of hospital-acquired pneumonia, the EPR is a reliable estimation of the period prevalence rate in comparison with the paper-based patient record. False positives and false negatives countervail each other, so that the rate of 0.9 % from the EPR lies within the 95 %-confidence limits calculated for the true value of 1.32 % from the paper-based patient record. This supports the potential of the secondary usage of electronically available routine data, as recently pointed out by Einbinder and Bates [5]. They stated that "... the secondary use of clinical data - including structured documentation, unstructured narrative, laboratory results, and administrative data - may help drive measurement, improve quality, and help manage populations of patients, even when they are not face-to-face with a provider". The estimated true rate of 1.32 % lies in the expected range of 0.5 % - 1.5 % published for acute hospital care [12]. The rate seems stable over time, taking into account the results from Leu et al. from 1989 with a period prevalence rate of 0.86 % [13]. External reference values are not available at the moment, because national as well as international quality indicator programs focus on appropriate treatment of pneumonia in outpatient care (cf. [14]). Nevertheless, the quality of data present at the University Clinics of Essen is sufficient for external quality control of hospital-acquired pneumonia as well.

In contrary, a correctness of 64.3 % and a completeness of 42.9 % underpin, that the ICD-10-coded diagnoses in the EPR at the University Clinics of Essen are not qualified for the daily support of patient care. The literature shows similar results. For example, Massanari et al. published a rate of 57.05 % of recorded nosocomial infections [15]. It seems realistic for Germany, that only half of the complications are at the moment coded in the electronically available routine data. Own experiences with other adverse events at the University Clinics of Essen are comparable, as in the case of pressure ulcer [16] and fall (unpublished). We argue that the insufficient completeness of adverse events in the EPR is to some extent caused by the minor role of the EPR at the point of care. As long as the control of the clinical processes is supported mainly by the paper-based patient record, the electronic one will not reach the necessary attention and consequently will lack in data quality. Additionally, our results confirm the conclusion of Mikkelsen and Aasly [17] that "parallel use of electronic and paper based patient records has resulted in inconsistencies" and "documentation is missing in both". This is a late recognition of Burnum's warning in 1989 [18], who argued that "medical record information has become less reliable than ever before despite the electronic information revolution in medical care ..". Burnum presumed a decrease in confidentiality induced by the electronic storage of data. Mikkelsen and Aasly point to the need of a systematic life-cycle-management of the EPR and adequate data-quality procedures.

Especially for the comparison of clinical data for a single individual one must be aware of a couple of relevant methodological difficulties as uncertainty of clinical reasoning, ambiguities of the definition of hospital-acquired pneumonia, and a weak reliability in diagnosis coding with the ICD-10 [19]. But, substantial difficulties are not specifically relevant for the EPR, they are as well apparent in the paper-based patient record.

For the University Clinics of Essen, the EPR is a reliable source for the frequency of hospital-acquired pneumonia, but not suitable to drive the care for an individual patient, e. g. in accessing appropriate guidelines. This supports the notion that electronically available routine data could be used within quality control programs. But

as long as they are recorded in parallel to the paper-based patient record they could not be easily incorporated into the process of health care. Regarding hospital-acquired pneumonia, a more precise definition and a merging with other data sources as lab results and radiology reports will improve correctness and completeness for both issues.

References

- [1] Ewig S, Dalhoff K, Lorenz J et al. Empfehlungen zur Diagnostik der nosokomialen Pneumonie. *Pneumologie* 1999; 53: 499-510.
- [2] Celis R, Torres A, Gatell JM, Almeda M, Rodriguez-Roisin R, Agusti-Vidal A. Nosocomial pneumonia. A multivariate analysis of risk and prognosis. *Chest* 1988; 93(2): 318-324.
- [3] Haley RW, Culver DH, White JW, Morgan WM, Emori TG, Munn VP, Hooton TM. The efficacy of infection surveillance and control programs in preventing nosocomial infections in US-hospitals. *Am J Epidemiol* 1985; 121: 182-205.
- [4] Raftery J, Roderick P, Stevens A. Potential use of routine databases in health technology assessment. *Health Technology Assessment* 2005; 9 (20).
- [5] Einbinder JS, Bates DW. Leveraging Information Technology to Improve Quality and Safety. *Methods Inf Med* 2007; 46 Suppl 1: 22-29.
- [6] Stausberg J, Kolke O, Albrecht K. Using a Computer-based Patient Record for Quality Management in Surgery. In: Cesnik B, McCray AT, Scherrer J-R, eds. *MedInfo 98*. Amsterdam: IOS, 1998: 80-4.
- [7] Bates DW, Evans RS, Murff H, Stetson PD, Pizziferri L, Hripscak G. Policy and the future of adverse event detection using information technology. *J Am Med Inform Assoc.* 2003; 10: 226-228.
- [8] Stausberg J, Koch D, Ingenerf J, Betzler M. Comparing paper-based with electronic patient records: lessons learned during a study on diagnosis and procedure codes. *Journal of the American Medical Informatics Association* 2003; 10: 470-477.
- [9] Stausberg J, Bartoszek G, Lottko B, Kröger K, Niebel W, Schneider H, Maier I. Problem focused Integration of Information, Quality and Process Management with Empirical Research: the Example of the Essen Interdisciplinary Pressure Ulcer Project. In: Park H-A, Murray P, Delaney C, eds. *Consumer-Centered Computer-Supported Care for Healthy People*. Proceedings of NI2006. Amsterdam: IOS, 2006: 609-612.
- [10] Garner JS, Jarvis WR, Emori TG, Horan TC, Hughes JM. CDC definitions for nosocomial infections. *Am J Infect Control* 1988; 16: 128-140.
- [11] Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. *J Am Med Inform Assoc.* 1997; 4: 342-355.
- [12] Ewig S, Ostendorf U, Müller E. Nosocomial pneumonia: diagnostics. *Dtsch Med Wochenschr* 2006; 131: 2661-2664 [in German].
- [13] Leu HS, Kaiser DL, Mori M, Woolson RF, Wenzel RP. Hospital-acquired pneumonia. Attributable mortality and morbidity. *Am J Epidemiol* 1989; 129: 1258-1267.
- [14] Agency for Healthcare Research and Quality. *Guide to Prevention Quality Indicators: Hospital Admission for Ambulatory Care Sensitive Conditions*. Version 3.1 (March 12, 2007).
- [15] Massanari RM, Wilkerson K, Streed SA, Hierholzer WJ. Reliability of reporting nosocomial infections in the discharge abstract and implications for receipt of revenues under prospective reimbursement. *Am J Public Health* 1987; 77: 561-564.
- [16] Stausberg J, Kröger K, Maier I, Niebel W, Schneider S, für das interdisziplinäre Dekubitusprojekt. Frequency of decubitus ulcer in patients of a university medical center. Combination of routine documentation and cross-sectional study. *DMW* 2005; 130: 2311-2315 [in German].
- [17] Mikkelsen G, Aasly J. Concordance of information in parallel electronic and paper based patient record. *Int J Med Inform.* 2001; 63: 123-131.
- [18] Barnum JF. The mininformation era: The fall of the medical record. *Annals of Internal Medicine* 1989; 110: 482-484.
- [19] Stausberg J, Lehmann N, Kaczmarek D, Stein M. Reliability of diagnoses coding with ICD-10. *International Journal of Medical Informatics* 2008; 77: 50-57.

Reliable Personal Health Records

Ton VAN DEURSEN ^{a,b}, Paul KOSTER ^a and Milan PETKOVIĆ ^a
^a Philips Research Eindhoven, The Netherlands
^b Université du Luxembourg, Luxembourg

Abstract. A number of applications based on personal health records (PHRs) are emerging in the field of health care and wellness. PHRs empower patients by giving them control over their health data. Health data for PHRs can be supplied by patients, wellness providers and health care providers. Health care providers may use the PHRs to provide medical care. Unfortunately, the quality of the health data in PHRs cannot be guaranteed in all cases. For example, consider cases where non-professionals such as patients and wellness providers supply data. To address this problem, we present in this paper a system that provides health care professionals with an indication of the quality of health data in a PHR. This indication is based on the reputation of the supplier and on metadata provided by measurement devices. The proposed reputation system mimics the way in which trust in health data and their suppliers is built in the real world. The system introduces minimal overhead for health care providers and patients.

Keywords. Quality management, EPR-CPR-EMR, telehealth, security

Introduction

For years, patient information has been maintained in paper-based records. Paper-based systems bring many disadvantages [1]. They have to be carried or faxed from one place to another, resulting in lack of availability of the records. Furthermore, illegible handwriting in paper-based records may lead to loss or misunderstanding of important health data [2]. Because of these disadvantages, paper-based systems are being replaced by electronic health record (EHR) systems. These EHRs are maintained by health care providers. Next to these EHR systems, personal health record (PHR) systems are developed [3,4,5]. The PHR is a health record maintained by the patient instead of the health care provider. PHRs empower patients by providing them control over their health data. The patient can manage and share his health data in his PHR at his own discretion. After sharing, the health data in the PHR can be used by health care providers and wellness providers to improve the patient's health.

Health care providers are most often the only parties providing health data for EHRs. However, health care providers are not the only parties that can provide health data for the patient's well-being. Patients (as well as people that are not ill, but are concerned about their health) may want to collect health data for their health records. Think for example of weight, heart rate and blood pressure information collected by the patient. Furthermore, as wellness providers such as fitness clubs and weight control clubs are professionalizing, they may want to use and provide relevant health data for

the patient's health record (see [4] as an example). This data is valuable information that can help a health care provider when treating the patient. The need of health care providers to use PHR data has been recognized by several standardization organizations that are already developing standards to enable exchange of information between EHRs and PHRs (e.g. IHE XPHR [6] and Continua xHR interface [7]).

For health data in a PHR, it is unknown how accurate and trustworthy this data is. This is because health data in a PHR is different from health data in an EHR. Health data in an EHR is always supplied (or at least reviewed) by a professional and may therefore be considered accurate and trustworthy. In contrast, health data in a PHR can be supplied by professionals as well as non-professionals. To be able to use the health data in a PHR, health care providers need to know whether this data is trustworthy and accurate.

The problem of assessing the reliability of websites is related to the problem of assessing the quality of PHR data. Assessing reliability of websites is done by measuring characteristics as completeness, accuracy and consistency. However, as measuring the quality of the measurement process for health data is very complicated we need a more general measurement for quality.

In this paper we propose Hedaquin, a system that provides a quality indication on health data in personal health records. This quality indication is based on the reputation of the supplier of the health data and on metadata provided by medical devices.

1. Approach

In order to have an indication of the quality of health data in a PHR, we design a reputation system that mimics the way in which trust in health data and their suppliers is built in the real world. To determine the quality of health data, the reputation of the supplier is used as a quality indication. In our approach, to determine the reputation we consider the following factors:

Credentials of the supplier: persons who have successfully completed a course of study are issued a diploma or certificate for that. Similarly, a credential may serve as a proof of quality or competence for a certain task. Diplomas, certificates and credentials are supplied by a trusted authority that is entitled to do so. For example, a person that has an academic degree in medicine has successfully completed a medical school. Therefore, the person can be expected to provide high quality health data.

The quality of previously supplied data: in case a non-professional supplies health data, a health care provider can state a judgment (i.e. give a rating) about the quality of the health data. The health care provider can do this if he observed the measurement process or if he verified the health data. The rating is later used in calculating the supplier's reputation.

Metadata: aside from the reputation, valuable information about the quality of health data can be supplied by medical devices. Medical devices can provide two types of metadata (information about the data supplied by the device). The first type of metadata is metadata on how the person uses the device (e.g. whether the device was used in a proper way). This type of metadata can be used to determine a rating that can be used in reputation calculation. The second type of metadata is metadata concerning the device itself (e.g. device model or calibration data). This type of metadata can be used to determine a quality indication of the data, but it does not influence the reputation of a supplier.

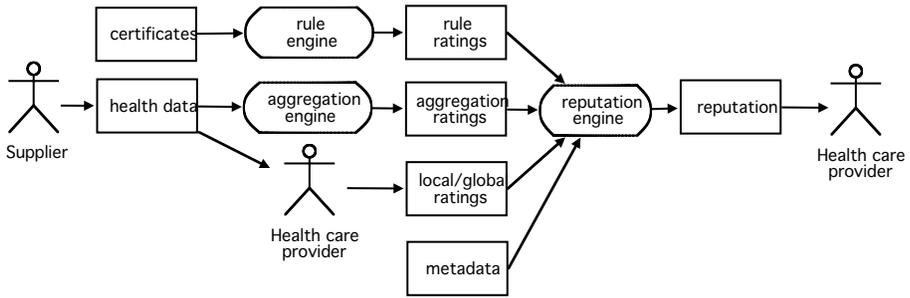


Figure 1. Hedaquin architecture

2. System Design

Figure 1 depicts the architecture of Hedaquin. The reputation engine takes ratings (local, global, aggregation and rule ratings) as input and calculates a reputation for a health data supplier, based on these ratings. Each rating is accompanied by a certainty level. The reputation engine is constructed in such a way that ratings with low certainty are given less weight than ratings with high certainty.

A reputation is a judgment on a supplier based on his past behavior. To represent a reputation, we use Jøsang's belief model [8]. We model a reputation as an opinion of a health care provider about a health data supplier. An opinion is a tuple (b, d, u) where b represents the belief in a statement, d represents the disbelief in a statement and u represents the uncertainty of a statement. An example of such a statement can be 'The person supplies high quality health data'. The reputation of the supplier, as calculated by the reputation engine, is used by the health care provider as a quality indication on the health data.

The aggregation engine calculates ratings based on the comparison of measurements and the rule engine creates ratings based on certificates of the supplier.

2.1. The Reputation Engine

The reputation engine calculates the reputation of a supplier of health data using available ratings. Suppose a health care provider calculates the reputation of a health data supplier. We distinguish four kinds of ratings that can be used in the calculation of the reputation:

Local ratings: ratings the health care provider himself has given to a certain supplier.

Global ratings: ratings that other health care providers have given to a certain supplier. There is a subtle difference between global ratings and local ratings. Because global ratings were not given by the health care provider himself, the certainty associated to these ratings must be decreased when calculating reputation. After all, a health care provider has less trust in other health care providers than in himself.

Aggregation ratings: there are some cases in which the quality of health data can be determined by comparing different sets of health data. If this is the case, a rating can be calculated automatically.

Rule ratings: a rule rating is based on the supplier's possession of certificates issued by independent organizations. Rule ratings are determined by the rule engine.

In order to calculate the reputation, we do not only consider the ratings themselves. We use additional information that is available on the rating and on the health data. This additional information determines the weight that is given to a particular rating. The following additional information is used:

The certainty of the rating: a rating with a high certainty should be given more weight than a rating with a low certainty. The certainty is supplied by the health care provider and indicates to which level he is certain about his rating.

The time of a rating: as the time between the creation of the health data and the calculation of the reputation part increases, the rating should be given less weight. After all, if a health data supplier performed good (or bad) a very long time ago, there is no guarantee that he will do so now. Wixted suggests that humans forget memories according to a power function [9]. Therefore, a power function should be used to mathematically capture the fact that reputation decays over time.

The order of the ratings: health data suppliers may start behaving better (or worse) over time. Therefore, recent ratings should be given more weight than older ratings. The last rating should be given the highest weight, the one before slightly less weight, etc. The weight assignment is done using a similar function as for the time of the rating.

The similarity in scopes: a scope relates to the task a health data supplier can perform, e.g. taking blood pressure measurements. If not enough ratings are available for a certain scope, ratings for similar scopes can be used. We use a scope function to assign high weights to ratings for similar scopes and low weights to ratings for dissimilar scopes.

Due to space limitations, the model and supporting formulas are not presented in this paper. We refer the reader to [10] for more technical details.

2.2. The Aggregation Engine

The aggregation engine automatically calculates aggregation ratings that are used to compute reputation. An aggregation rating is computed by comparing measurements from different suppliers with a small time difference. If two health data suppliers (e.g. a doctor and a patient) take the same measurement on the same person and these measurements are similar, then the reputation of both suppliers can be increased. If two suppliers take the same measurement on the same person and the measurements are not similar, then the reputation of both suppliers must be decreased. The amount with which the reputation should be increased or decreased depends on the reputation of the suppliers that take the measurements. Therefore, in the case of a doctor and a patient taking the measurement, the reputation of the patient is much more influenced than the reputation of the doctor.

2.3. The Rule Engine

The rule engine determines rule ratings that can be used by the reputation engine. These rule ratings are determined based on the certificates the supplier for which the reputation is calculated possesses. The rule engine uses a predefined mapping to find the rule ratings associated with a certificate. Examples of certificates are medical diplomas or certificates for completing courses.

3.2. Integration in Clinical Client Software

While entering numerical values is an easy and convenient way to provide ratings, health care providers may prefer a graphical method for entering ratings. A simple solution is providing the ratings by means of sliders (Figure 3). The value of the ratings is entered by sliding to the right (for a high rating) and to the left (for a low rating). The slider for entering the certainty of the rating can be used in a similar way.

Recall that a reputation is represented by an opinion of the form (b,d,u) . The advantage of using an opinion is that opinions are normalized (i.e. $b+d+u=1$). Therefore, we can represent an opinion by a bar where the values of b,d and u determine the look of the bar. The belief b determines the size of the green part of the bar, the disbelief d determines the size of the red part of the bar and the uncertainty u determines the white part of the bar (Figure 4). By visualizing the reputation instead of showing numerical values, the quality indication is much faster and easier to comprehend.

4. Conclusion

In this paper, we proposed to use a reputation system and metadata provided by measurement devices to give a quality indication for health data in personal health records. Therefore, health care providers can use health data supplied by non-professionals, such as patients and wellness providers, in a much safer way. This also results in reduced costs and higher quality of health care.

The purpose of a reputation system is to build trust in online environments and to provide an incentive for good behavior. Therefore, using a reputation system to make a quality indication of health data is a natural choice. It also mirrors real practice, in which a health care provider builds trust in patients and the health data they supply.

All in all, Hedaquin gives health care providers the opportunity to make an informed decision on the quality of health data that is supplied by patients and wellness providers. For patients and wellness providers there is no overhead in using the system.

Future research will concentrate on testing the concept in a medical setting and on the use of metadata to determine health data quality and ratings for health data quality.

References

- [1] Commission on Systemic Interoperability. Ending the Document Game. U.S. Government Printing Office 2005.
- [2] Kohn L, Corrigan J, Donaldson M. To err is human: Building a safer health system. The National Academies Press 2000.
- [3] WebMD HealthManager. <http://healthmanager.webmd.com> (February 15, 2008).
- [4] LifeSensor. <http://www.lifesensor.com> (February 15, 2008).
- [5] Dossia. <http://www.dossia.org> (February 15, 2008).
- [6] Boone K. Exchange of personal health record content (XPHR). 2006.
- [7] Continua EHR interface. <http://www.continuaalliance.org> (February 15, 2008).
- [8] Jøsang A. A logic for uncertain probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 2001; 9(3):279-311.
- [9] Wixted JT. On Common Ground: Jost's (1897) law of forgetting and Ribot's (1881) law of retrograde amnesia. *Psychol Rev.* 2004; 111(4):864-79.
- [10] van Deursen T. Hedaquin: A reputation-based health data quality indicator for a personal health record system. Master's thesis. Technische Universiteit Eindhoven 2007.

Involving Clinicians in the Development of an Electronic Clinical Handover System – Thinking Systems Not Just Technology

Ming Chao WONG^{a,1}, Paul TURNER^a and Kwang Chien YEE^{a,c}

^a*School of Information Systems, University of Tasmania, Australia*

^c*Royal Hobart Hospital, Tasmania, Australia*

Abstract While the need to ‘involve the user’ in information technology (IT) development is almost a mantra amongst information systems specialists, numerous IT projects continue to fail because of an inability to capture user insights or respond to users needs. Although there are clearly practical difficulties in addressing and responding to the heterogeneous requirements expressed by different users, marginalizing these views ultimately is to the detriment of the systems built. This paper describes the development of an electronic clinical handover system at the Department of General Internal Medicine (DGIM), Royal Hobart Hospital (RHH). More specifically, the paper aims to highlight how to engage meaningfully with clinicians in the development of a sustainable system. It is anticipated that by drawing attention to the importance of users and by outlining the practical experience of dealing with the diversity of requirements and views expressed, the paper can contribute to a stronger recognition within the domain of eHealth for a user-centred systems approach to IT development.

Keywords. Socio-cultural systems, clinical handover, information communication technology, health informatics, patient safety

Introduction

It has previously been reported that 75% of large IT projects in health care fail [1]. This astonishingly high figure is at least partially attributable to a failure on the part of developers to understand the workflow of health professionals and to meaningfully involve users in the design, development and implementation of these systems [2].

Information Systems (IS) has traditionally bridged this divide between end-users and technology developers. As a discipline it is founded on recognition of the importance of user involvement in systems design, implementation and evaluation. In recent years, a variety of approaches have been developed to support user involvement including user-centred design; participatory design; prototyping and joint application design and human factor engineering. All of these approaches have been used to improve information systems analysis, design and implementation processes [3].

¹ Corresponding Author: Ming Chao Wong, School of Information Systems, University of Tasmania, Private bag 87, 7001, Hobart, Tasmania, Australia.; E-mail: mcwong@utas.edu.au

This paper describes the development of an electronic clinical handover system at DGIM at the RHH and describes the value of involving clinicians, particularly junior medical officers (JMOs) in all stages of system analysis and design. More specifically, the paper aims to outline the methodological approach of involving and empowering clinicians right from the start of its development. This approach aims to acquire an in-depth understanding of the clinical handover process and to ensure the enfranchisement in the development process.

1 Project Background

The RHH is Tasmania's main tertiary referral public hospital, operating within the acute care service division of the Tasmania Department of Health and Human Services (DHHS). The RHH utilises traditional project management methodologies which advocate the use of committees involving major stakeholders in decision making processes [4]. However, the traditionally hierarchical nature of medicine often means that if any clinicians are invited to be part of a committee they are usually senior consultants rather than actual users of the proposed system [5]. Unfortunately, as in many other jurisdictions, this can result in JMOs who are often in the frontline with regard to the use of hospital IT systems rarely being asked for direct input or feedback on the system [6]. As research has revealed these approaches often lead to problems when systems are implemented as they fail to fit into existing workflow practices of JMOs. This often becomes a hindrance to clinical routine [7].

While ICT developers are often keen to build state-of-the-art systems, utilising cutting edge technologies, with multiple user interfaces, wireless accessibility and high levels of system functionality, experience suggests that the utility of these systems for clinicians' daily practice remains debatable. Although system functionality and feature complexity is of interest to designers, it is imperative in the health domain that there is an acknowledgement that many end-users continue to have difficulty in incorporating basic information systems functionality into their activities. This suggests not only a need for training but also for user involvement at all stages of the system development life-cycle. Critically, there is a need for more practical examples of how to support this user involvement amongst clinicians.

This paper examines these issues in the context of the development of an electronic clinical handover tool and illustrates how JMOs and senior management were both actively involved in the decision-making processes. The paper argues that this approach provides significant benefits for implementation of the system.

2 Methodology

2.1 Stakeholder endorsement

This project commenced by acquiring endorsement from all stakeholders involved to ensure support at all levels. More specifically, this meant endorsements from DGIM of RHH, School of Medicine and School of Information Systems, University of Tasmania, as well as DHHS, Tasmania State Government.

At a practical level, this involved researchers interacting closely with all levels of clinicians at the hospital from interns to consultants via observations, interviews and focus groups in an iterative manner. All levels of clinicians were invited to participate in the early stages of the project regardless of their clinical seniority. While the focus was on JMOs as the primary end-users, comments, insights and experiences were also captured from senior clinicians. These processes ensured that user requirements were identified and that system functionality could be trialled and tested directly with users.

The information systems (IS) researchers were not medically trained and initially had to deal with resistance from some clinicians who initially perceived the project as an effort to introduce technology for 'policing' their clinical practice. There were also some concerns expressed that an IS researcher would not understand clinician practice due to a lack of medical knowledge. As a consequence of this initial resistance there was a need for time and energy to be expended in building rapport with many of the JMOs through informal chats and attending informal gatherings and social activities. This resulted in building trust between clinicians and the research team – a critical factor in ensuring that user requirements could be captured accurately.

2.2 Familiarisation with Clinical handover

The research team utilized two main data collection techniques in the process of familiarisation with the actual practice of clinical handover. Firstly, 30 observation sessions were carried out during the morning handover and night handover over three extended holiday periods where staff numbers were down to their minimum and adequate handover was extremely important for patient care. Field notes were obtained and typed into the computer at the end of the observation sessions independently by an information systems researcher and a clinician. Secondly, semi-structured interviews were conducted individually with four groups of clinicians: experienced interns, inexperienced interns, registrars, and consultants.

The interviews were audio-recorded and transcribed. Analysis of the transcripts was conducted using open, axial and selective coding drawing on the principles of grounded theory [8]. The data from the observations and interviews were then reconciled to obtain an in-depth understanding of the clinical handover process. This in-depth understanding was then again discussed with clinicians at the hospital.

2.3 Systems Development

The first step in determining the requirements for systems development involved an analysis of 50 handwritten clinical handover messages provided by the clinical handover leader. This was done to identify the minimal field data set required for effective clinical handover.

A series of 6 workshops drawing on the principles of participatory design [9] were then conducted to work through the minimal data set and IT specifications. The workshops involved clinicians of all levels but were aimed primarily at JMOs as they were to be the main users of the system. Clinical examples were used and opinions from participants regarding the data requirements, presentation and delivery were documented. The documented opinions from the clinicians were then analysed and a simple template drawn up which was representative of the final IT system. This template was then distributed to end-users in both electronic form and printed form for comments and further discussion and analysis.

3 Results and Discussion

3.1 “Mythical” vs Actual clinical handover process

The analysis of the data collected through observations and semi-structured interviews revealed that there was a significant difference between what was thought to be the handover process and what actually happened at handover. Interviews conducted with consultants revealed that clinical handover was an efficient process with a well formulated structure facilitating adequate information exchange and was attended by all involved and at times included an education component for JMOs. Observations sessions however, revealed that there were multiple factors, including cultural, environmental, human and other factors that determine the structure, attendance and efficiency of the clinical handover process. Furthermore, apart from the primary function of information exchange, there were other functions embedded within the clinical handover process eg. Seeking a second opinion, a reminder to check test results, seeking supervision and debriefing of difficult and stressful situations.

Our experience of identifying a “mythical” and an actual clinical handover process closely mirrored the description of the problem explained by Norman [5]. Domain experts have little insight into their own work and they are often divorced from actual use [5]. Our study has validated Norman’s argument in the clinical handover setting that observations by trained professionals are important in understanding workflow.

3.2 Building the electronic clinical handover system

Given the difference between the “mythical” handover process and the actual handover process, a decision had to be made as to whether the system was built to address the needs of the “mythical” handover or the actual handover process. A decision to build an IT system to support the “mythical” handover process would risk failure because it would not fit in with what was actually done at handover. A decision to build an IT system to support the actual handover process would risk failure because some clinicians might feel that their opinion was not taken into consideration.

Following further discussion, it was agreed that a system be built to support the actual work processes as this had the strongest support and the best likelihood of adoption by JMOs. The IT project team understood fully the different approaches utilized in clinical handover by different clinicians in actual workflow. The formalisation of the clinical handover process through IT program implementation trajectory will only be productive if this is aligned with the organizational change trajectory [2]. Therefore, during the implementation process, educational and leadership program for clinical handover will be implemented simultaneously.

3.3 Data field determination

Once the decision is made to build a web-based IT system to support clinical handover, data fields needed to be determined. In moving forward with the development, the written clinical handovers notes were analysed. This provided the project team with the following minimal data fields for clinical handover program: 1. Patient unique

identifier number, 2. Patient's demographic details, 3. Location and treating team, 4. Background issues and 5. Action required.

Minimal data fields for clinical handover had been published by others recently [10], [11]. Bernau et al provided an expert opinion on the minimal field requirement [11], while Cheah et al identified minimal data fields for surgical handover but did not mention the method utilised to identify these data [10]. We found that more data fields were required to provide adequate clinical handover compared to those expressed by Bernau et al [11]. We further identified background issues and actions required as separate fields while Cheah et al. grouped them into free-text [10]. In our opinion, the separation of issues and actions required is important and we see the implementation of the electronic handover system as a good opportunity to drive the changes to separate the two data fields to improve the effectiveness of clinical handover.

A series of workshops, with emphasis on the primary users – JMOs were then conducted. These workshops not only reconfirmed the minimal field data, but also addressed the issues of delivery, information presentation, resource requirements as well as additional features – automated display of pathology results, to support the workload of junior staff. While we acknowledge that having greater clinician participation might not help [2], we believe that having the right clinician participation, i.e. the primary users of the system will help not only with the design, but also with the feeling of ownership of the program.

3.4 *Safety feature design*

In order to improve the clinical handover process and to improve patient safety, we added a few in-built safety features with the electronic system. These safety features have been described elsewhere [12]. The emphasis on safety is essential. More importantly, we believe that a clear communication of the safety features of the system is important to improve sustainability. Clinicians are intrinsically driven by the virtue of “saving lives”. By aligning the strategic aims of the electronic clinical handover to both organization change trajectory and personal intrinsic desire of clinicians, we believe that it forms a force of empowerment that will drive sustainability.

3.5 *Create motivation and sustainability*

While the researchers acquired numerous comments, insights and recommendations from the JMOs regarding the features they wanted out of the electronic clinical handover system, it was noticeable that these were very different as compared with those acquired from senior clinicians about the system. JMOs emerged as eager to utilize the system and believed that the system would provide genuine benefits. They also indicated their preferences for complicated technologies, such as bar-coding for user identification and wireless mobile technologies. On the flip side, most senior clinicians indicated their preference for forms that could be printed out so that they can make their own hand-written notes on them. That internal conflict is balanced out by taking a middle ground by the research team, communicating clearly that a plan for version management is in place if the uptake of the program is encouraging. We believe that this approach creates an incentive for sustainability.

The research team also identified proponents and opponents for system changes and considered this clear dichotomy of views regarding the IT system and the value of process improvement. The research team included both these perspectives into the

design phase and opened up discussions on these issues to acquire further feedback. Noticeably, it was revealed that opponents of the process improvement actually provided some of the most valuable input into the design specification.

The researchers utilized the opportunistic reflective learning as a method to empower clinicians and to create an environment of readiness for change. By providing clinical examples in the workshop, clinicians could better relate to the need to improve the process and the potential benefits of this IT project. Additionally incentives for clinicians to use the system were also included. These incentives were identified through interviews and observations and built into the IT design. We believe that the above methods of education and training, listening through group workshops, obtaining feedback and good communication, version management and utilization of self-reflective learning will empower clinicians for IT sustainability in healthcare.

4 Conclusion

This paper has outlined the mechanisms used by IS developers to involve clinicians in the process of developing an electronic clinical handover system. It has examined the practical decisions that have been made in dealing with heterogeneous inputs from the interactions with a diverse range of perspectives and attitudes towards the system to be developed. The paper reveals that in working towards the development of a sustainable system it is important to engage all stakeholders, obtain first hand data (i.e observations), involve clinicians throughout in the development phase, align the IT project trajectory to organisation change trajectory and provide clinicians with opportunity for self-reflection and self-actualisation.

References

- [1] Littlejohns P, Wyatt JN and Garvican L. Evaluating computerised health information systems: hard lessons still to be learnt. *BMJ*. 2003; 326: 860-863.
- [2] Wears RL and Berg M. Computer technology and clinical work: Still waiting for godot. *JAMA*. 2005; 293: 1261-1263.
- [3] Valacich JS, George JF and Hoffer JA. *Essentials of Systems Analysis and Design*. 2nd ed. New Jersey: Pearson Education Inc. 2004.
- [4] Tasmanian State Government. Project Management available [<http://www.projectmanagement.tas.gov.au>] accessed 3 December 2006.
- [5] Norman DA. *Emotional Design: Why We Love (or Hate) Everyday Things.*: NY: Basic Books. 2004.
- [6] Berg M. *Health Information Management: Integrating In-formation Technology in Health Care Work*. London, Eng-land: Routledge. 2004.
- [7] Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE and Strom BL. Role of computerized physician order entry sys. in facilitating medic. errors. *JAMA*. 2005; 293: 1197-1203.
- [8] Neuman WL. *Social Research Methods: Qualitative and Quantitative Approaches*. 3rd ed. Boston: Allyn & Bacon. 1997.
- [9] Bodker K, Kensing F and Simonsen J. *Participatory IT Design: Designing for Business and Workplace Realities*. Cambridge: The Massachusetts Institute of Technology Press. 2004.
- [10] Cheah L, Amott DH, Pollard J and Watters DAK. Electronic medical handover: towards safer medical care. *Med J Aust*. 2005; 183: 369-372.
- [11] Bernau S, Aldington S, Robinson G and Beasley R. From medical student to junior doctor: the medical handover – a good habit to cultivate. *StudentBMJ*. 2006; 14: 188-189.
- [12] Wong MC, Turner P & Yee KC. Socio-Cultural Issues and Patient Safety: A Case Study into the Development of an Electronic Support Tool for Clinical Handover. *Stud Health Technol Inform*. 2007; 130: 279-289.

Perfect Match? Generation Y as Change Agents for Information Communication Technology Implementation in Healthcare

Kwang Chien YEE^{a,b,1}, Erin MILLS^{a,b} and Caroline AIREY^c

^aRoyal Hobart Hospital, Tasmania, Australia.

^beHealth Services Group, School of Information Systems, University of Tasmania.

^cRoyal Brisbane Hospital, Queensland, Australia.

Abstract. The current healthcare delivery model will not meet future healthcare demands. The only sustainable healthcare future is one that best leverages advances in technology to improve productivity and efficiency. Information communication technology (ICT) has, therefore, been touted as the panacea of future healthcare challenges. Many ICT projects in healthcare, however, fail to deliver on their promises to transform the healthcare system. From a technologist's perspective, this is often due to the lack of socio-technical consideration. From a socio-cultural perspective, however, there is often strong inertia to change. While the utilisation of user-centred design principles will generate a new wave of enthusiasm among technologists, this has to be matched with socio-cultural changes within the healthcare system. Generation Y healthcare workers might be the socio-cultural factor required, in combination with new technology, to transform the healthcare system. Generation Y has generated significant technology-driven changes in many other industries. The socio-cultural understanding of generation Y healthcare workers is essential to guide the design and implementation of ICT solutions for a sustainable healthcare future. This paper presents the initial analysis of our qualitative study which aims to generate in-depth conceptual insights of generation Y healthcare workers and their view of ICT in healthcare. Our results show that generation Y healthcare workers might assist future ICT implementation in healthcare. This paper, however, argues that significant changes to the current healthcare organisation will be required in order to unleash the full potential of generation Y workers and ICT implementation. Finally, this paper presents some strategies to empower generation Y workers as change agents for a sustainable future healthcare system.

Keywords. Socio-technical integration, qualitative research, organisational culture

Introduction

The healthcare system is at a cross-road. Advances in technology over the last few decades have improved the delivery of care and significantly improved life expectancy.

¹ Corresponding Author: Kwang Chien Yee, School of Information Systems, University of Tasmania, Private bag 87, 7001, Hobart, Tasmania, Australia.; E-mail: kwang.yee@dhhs.tas.gov.au

This legacy, however, is now the biggest challenge in maintaining a sustainable, high quality healthcare system. Our patient population is ageing rapidly and the care deliveries of patients with multiple co-morbidities have rapidly increased in complexity [1]. In the coming decades, the healthcare system will have to undergo the biggest transformation of any industry in history in order to deliver best quality care to the community [2]. Furthermore, the increasing emphasis and utilisation of multidisciplinary care creates the need to improve communications and information exchanges among various healthcare professionals [3]. As resources to deliver care are not unlimited, the current healthcare delivery model will not be sustainable in the near future. The only sustainable future for healthcare delivery is one that best leverages technology to improve productivity and efficiency.

ICT implementation in healthcare, however, has not so far delivered on its promise to improve efficiency and safety of healthcare delivery [4]. Significantly, many information technology projects in healthcare fail [5]. When the failures of ICT are analysed in detail, it becomes clear that the problem is not related to the technology itself, but the lack of socio-technical consideration [6]. There has been a recent proliferation of various user-centred design techniques which claim to achieve better socio-technical integration [6] [7]. Future emerging technologies, armed with socio-technical integrated design, will likely lead to a level of renewed enthusiasm in the electronic transformation of future healthcare [8].

From the socio-cultural perspective, there is an equally powerful force which will transform the healthcare system: the entry of generation Y clinicians. Generation Y is commonly defined as those born after 1978 [9]. This generation of young workers has created big impacts in other industries [10]. Their entry and the impact of their entry on the healthcare system are slowly being acknowledged. Their familiarity with technology is likely to influence healthcare organisation to embrace the digital future. Is generation Y the socio-cultural answer to the prayers of enthusiastic health information technologists? Will the combination of emerging technology and generation Y healthcare workers create the momentum necessary to drive us towards a sustainable healthcare future? It is therefore essential and timely for us to investigate the impact of generation Y workers on ICT implementation in healthcare.

The paper firstly describes our qualitative methodological approach to generate in-depth understanding of generation Y healthcare workers. The results to date are then presented. This paper presents the argument that generation Y phenomenon is the golden opportunity to propel electronic transformation of the healthcare system. There are, however, many challenges that healthcare systems will need to face in order to best leverage the potential of generation Y workers. This paper aims to contribute significantly to the health informatics field by presenting some potential solutions through the lens of generation Y to empower this generation of healthcare workers to contribute to the sustainable digital future healthcare.

1. Methodological Approach

Given the new socio-cultural experience created by generation Y in other industries [9], it is imperative that the healthcare system takes generation Y seriously. Firstly, the future design and implementation of ICT within the healthcare system needs to take generation Y into account. Secondly, and more importantly, as generation Y slowly

replaces the ageing healthcare workforce, an in-depth socio-cultural understanding of generation Y is required in order to create a sustainable healthcare future.

Given the scanty literature available on this topic, the main aim of this pilot exploratory study is to generate in-depth socio-cultural understanding of generation Y doctors within the current healthcare setting. This research project has the following objectives:

- To generate an understanding of the impact of generation Y on healthcare.
- To understand the perception of ICT in healthcare among generation Y-ers.
- To generate guiding principles to involve generation Y as change agents for ICT implementation in healthcare.

In order to obtain a holistic understanding regarding the socio-cultural issues of generation Y, we deploy qualitative research methodology, applying interpretivist epistemology as our research philosophy. We use a three-stage approach in order to generate complex socio-cultural contextual insights. Given the technological orientation of generation Y, the researchers take this opportunity to explore the possibility of undertaking technologically-supported qualitative research. The data recorded are analysed using Giorgi's descriptive phenomenological method [11].

The detail description and justification of our research methodological approach has been published elsewhere [12]. Multiple steps have been built into our methodological approach in order to answer the call for improved rigors of qualitative research [12]. A brief description of our research methodology is provided here.

1.1. First Stage; Familiarisation Phase

This phase of the research carried out in 2006, involved the utilisation of ethnography and semi-structured interviews techniques. One of us had the privilege to work with 15 generation Y doctors and especially had the privilege to work closely with five of these doctors at different period of time as a team to deliver healthcare services. These five doctors became the key informants and the other ten doctors became informants of this phase of study, at which social interactions and interviews were documented; especially when it was perceived that these generation Y doctors exhibit certain behaviours to display their cultural believes. The "thick description" method described by Clifford Geertz was utilised in the event-triggered documentation [13]. These field notes, taken over the 12 months period in 2006, were then analysed using Giorgi's descriptive phenomenological method [11].

The researcher then carried out ten semi-structured interviews with final year medical students. Field notes were taken during the interviews. The interviews were recorded if the interviewees agreed to the process. The transcript of the interviews were analysed using the Giorgi's descriptive phenomenology methodology [11].

1.2. Second Stage: Longitudinal Semi-Structured Email Interviews

The second phase of the study involves a longitudinal follow up study of generation Y doctors in 2007. All generation Y interns at the Royal Hobart Hospital who are willing to participate in the study were sent weekly emails for the first month and then monthly emails for 12 months. Each email contained five open-ended questions as essential questions. The first response to the email will be screened for the need for clarifications and probing. The analysis of the answers will be carried out within three days and

clarifications will be sought from responders if needed. The data collected will be coded and analysed using Giorgi's descriptive phenomenological method, by generation Y doctors [11].

1.3. Third Stage: Traditional Interviews at Two Separate Sites

Phase three of the study will be carried out at two separate sites late 2007. All generation Y interns at the Royal Hobart Hospital will be invited to participate in a face-to-face interview. These interviews will be audio recorded and transcribed. This same process will be carried out at a different site to a different cohort of generation Y medical doctors by a separate generation Y researcher. The same research questions and techniques will be used. The data will be coded, analysed and interpreted by the two researchers separately.

2. Results

We have completed the phase 1 study and in the process of carrying out phase 2 and phase 3 of the study. This paper will present results from phase 1 and part of phase 2 data. Our results identify five key characteristics of generation Y doctors, which will affect future ICT design and implementation in healthcare.

Firstly, generation Y doctors are technology savvy. Generation Y doctors are not only familiar with technology but also have strong belief in their ability to rapidly adapt to technology. They, however, prefer to use technology that they are familiar with, rather than cutting edge technology. Generation Y doctors welcome new technology implementation as it provides the necessary stimulation and challenges that are vital to the perception of job satisfaction for generation Y. Secondly, generation Y doctors gather information on a 'PRN' basis – they want on-demand information and education relevant to what they are doing right now. Electronic means form the most common way of information gathering and obtaining decision support. Google is included in the list of resources that junior doctors learned the most from during their internship year! Online resources provide the bulk of resources used in everyday work and for ongoing education. The expectation of gen. Y doctors is that these resources are made available to them by healthcare organizations. They want these resources to have efficient search engines to provide fast, reliable, relevant, sufficiently detailed and up-to-date information. In generation Y terms: "U get info U wnt 2 ur mates only. :)".

Thirdly, our results suggest that one of the top priorities for Generation Y doctors is a well-balanced lifestyle with plenty of life experience. They thrive in environments that are highly stimulating and those which will challenge them. Generation Y doctors are likely to move regularly from one institution to the other in search for better education, training and life experience. While they acknowledge the current ICT systems are different at different institutions, they do not believe that they need training to adapt to new ICT. They believe that the system design should be "user-friendly" enough for them to adapt to without training.

Fourthly, generation Y doctors have very high expectations for ICT in healthcare. They expect that new ICTs continue to be designed and implemented to improve workplace efficiency and to reduce their workload. Furthermore, Generation Y doctors expect well-designed ICTs to be available for utilisation in their working environment;

in their terms “user-friendliness just like google, you don’t need to learn and yet you can find what you want.”

Finally, generation Y doctors expect the organisations they work for to provide them with adequate hardware and software. They expect desktop computers to be available in sufficient quantities and in appropriate locations, each with high-speed internet access. They also feel that healthcare institutions should provide them with personalized, portable computers such as Personal digital assistants (PDAs) with wireless technology. More importantly, despite their very high expectations and demands regarding ICT, Generation Y doctors dissociate themselves from the design and implementation of new ICT. They do not believe that they could participate and assist in the socio-technical integrated design of ICT, nor do they see it as their role.

3. Discussion

Generation Y healthcare professionals have many characteristics which makes them powerful change agents within the healthcare sector for ICT implementation. They are not only technology savvy, but they have great confidence in technology. Generation Yers are confident in their abilities to adapt to new technologies, and they welcome new challenges. Generation Y doctors will demand and expect organisations to provide them with well designed ICT solutions to support their work. This will generate significant momentum to fasten the pace of ICT adaptation and implementation within the healthcare system. As significant shortage of healthcare professionals continues to spread across the world, employers have to be creative in recruitment and retention of staffs. It seems likely that one of the strategies to recruit and retain generation Y doctors is the implementation of ICT to automate many routine tasks.

Generation Y will move from one institution to another in search of new experiences and a better lifestyle. The biggest challenge faced with different institutions is the acquisition of tacit knowledge to effectively fulfill their roles. The fluidity of generation Y workforce combined with their desire for technology will motivate institutions to implement better ICT supported organisational tacit knowledge management strategies.

The more important effect of generation Y in the healthcare, however, is the ripple effect generated by the combination of generation Y doctors and well designed technology. The enthusiasm of Generation Y doctors using new technologies to improve the effectiveness and efficiency of healthcare delivery will affect the perception of ICT implementation by senior clinicians. If this powerful combination is leveraged appropriately, a “bottom-up” approach to ICT adaptation will ensure.

The concept that generation Y professionals might be the change agents for ICT implementation in healthcare is very attractive to health informatics professionals. This concept, however, faces some steep challenges in real-world. Firstly, generation Yers have high expectations of potential benefits of technology. The current available technology might not fulfill the expectation of generation Y professionals. Furthermore, Generation Y professional cannot fulfill the role of effective change agents within the current healthcare organizational structure. The healthcare system is traditionally a hierarchical organization with senior members making critical decisions and the junior doctors providing the bulk of patient care. The technology gap and generation gap are widening between generation Y doctors and senior clinicians. More importantly,

organizations, with strong hierarchical tradition will likely to expect generation Y to conform to the culture rather than embrace new changes.

This might explain our observation that generation Y doctors are reluctant to participate and assist ICT design as they lack the autonomy and authority required to be heard. There is a strong need to re-engineer the healthcare hierarchical system in order to best leverage the potentials of generation Y workers. Generation Y workers need to be given the necessary autonomy and be involved in every step of the design and implementation process. In fact, some generation Y healthcare professionals should be encouraged to become leaders to assist ICT design and implementation. Our approach of utilising generation Y doctors as researchers in order to conduct and present research through the lens of generation Y highlights one of these strategies to empower generation Y healthcare professionals for a sustainable digital future!

4. Conclusion

In this paper, we presented our views of changing socio-cultural factors in healthcare. We presented our methodological approach to investigate this socio-cultural change: the entry of generation Y into the healthcare system. We presented our results to date on the potential socio-cultural impact of generation Y on ICT implementation in healthcare. We argued that generation Y workers carry with them many characteristics which allow them to be engaged as change agents for a sustainable digital future. There are many challenges in engaging generation Y in ICT implementation. The most important challenge of all is the hierarchical medical tradition, which prevents them to effectively fulfill that role. There is an urgent need to re-engineer the healthcare system in order to empower generation Y and listening to our view is a good start!

References

- [1] Yee KC, Barbour L. Improving medication safety at the interface: The role of pharmacist versus information technology in the real world. *Program and abstracts of Improving Medication Safety Conference: Sharing the lessons learned*, Sydney, New South Wales, Australia, 2006.
- [2] Fitz-Enz J. Industry change = internal change. HRD/OD, *Am. Management Assoc.Int.* **1** (1998), 5.
- [3] Yee KC, Showell C. Clinicians' perceptions of digital medical records: an Australian case study. *HIMSS Asia Pacific*. Singapore. 2007
- [4] Turner P, Wong MC, Yee KC. Medical Error Management: The Need for Information Systems Not Just Technology for Supporting Patient Safety in Hospitals. *Proceedings of the 14th Annual Health Informatics conference*. Melbourne: Health Informatics Society of Australia. 2005.
- [5] Littlejohns P, Wyatt JN and Garvican L. *Evaluating computerised health information systems: hard lessons still to be learnt*. *BMJ* **326** (2003), 860-863.
- [6] Wears RL, Berg M. Comp. techn. and clin. work: Still waiting for Godot. *JAMA* **293** (2005), 1261-1263.
- [7] Beuscart-Zephir MC, Elkin P, Pelayo S. *Human factors engineering for clinical application*, *Stud Health Technol Inform.* **124** (2006), 685-90.
- [8] Coiera E, *Putting the tech. back in socio-technical sys. res*, *Int J Med Inform* **76 Suppl 1** (2007), 98-103.
- [9] Sheahan P. *Gen. Y: Thriving and Surviving with Gen. Y at Work*. Hardie Grant Books, Victoria, 2005.
- [10] Salt B. *Y the younger generation is so different*. The Australian. 16th November (2006), p. 26.
- [11] Giorgi, A., & Giorgi, B.. *The Descriptive Phenomenological Psychological Method*. In P. M. Camic, J. E. Rhodes & L. Yardley (Eds.), *Qualitative Research in Psychology. Expanding Perspectives in Methodology and Design*. American Psychological Association. Washington. DC, 2003.
- [12] Mills E, Airey C & Yee KC. *Gen. Y in Healthcare: The Need for New Socio-technical Consideration for Future Technology Design in Healthcare*. *Stud Health Technol Inform* **130** (2007), 169-179.
- [13] Neuman WL. *Social Resh. Meth.: Qualitative and Quantitative Appro*. Allyn and Bacon, Bosteon, 1997.

This page intentionally left blank

6. Human-Computer Interaction & Imaging

This page intentionally left blank

Does a Hybrid Electronic-Paper Environment Impact on Health Professional Information Seeking?

E. M. BORYCKI RN PhD^{a,1} and L. LEMIEUX-CHARLES PhD^b

*^aSchool of Health Information Science, University of Victoria,
British Columbia, Canada*

*^bDepartment of Health Policy, Management and Evaluation, University of Toronto
Toronto, Ontario, Canada*

Abstract. The purpose of this study was to investigate the effects of a hybrid electronic-paper patient record environment upon health professional information seeking (i.e. amount of information accessed, choice of key sources of information, type of information and use of information seeking tactics). A within group, laboratory, experimental study was conducted using two simulation environments (i.e. a paper patient record and a hybrid or electronic-paper environment). Thirty-five novice nurses participated in this within group, laboratory based study. Findings revealed significant differences between the paper and hybrid environments in terms of their effects upon information seeking. Subjects: (1) accessed less data in the hybrid than the paper environment, (2) accessed more non-electronic sources than electronic sources of information in the hybrid environment, and (3) used more passive information seeking tactics in the hybrid than the paper environment. Findings from the cued recall data revealed subjects experienced increased cognitive load in the hybrid environment. Implications for the design of hybrid environments are discussed.

Keywords. EPR-CPR-EMR, Health professional workstation, human interfaces

Introduction

Health services and health informatics researchers are recognizing that the ability of clinicians to acquire, interpret, and use patient information is a critical aspect of clinician [1-3], team [5] and organizational decision making [4,5]. Yet, even as the process of information seeking remains a critical aspect of health care work [10], clinicians spend a considerable amount of their time seeking information [1,11]. Research suggests up to 70% of clinician information needs relevant to patient care decision making are unmet [1, 6]. Furthermore, up to 42% of workers accidentally use the wrong information, and 57% of workers go to many information sources to obtain needed information [4, p. 1]. Given the urgency associated with providing timely and

¹ Corresponding Author: E. M. Borycki, School of Health Information Science, University of Victoria, P. O. Box 3050 STN CSC, Victoria, British Columbia, V8W 3P5, Canada; E-mail: emb@uvic.ca

relevant information to clinicians, information seeking has become an important area of concern for health care organizations.

1. Literature Review

Health care organizations are investing in information technology (IT) to support clinician information seeking [2, 4]. A number of differing types of IT are being deployed in health care organizations to support clinician information seeking, among them the electronic patient record (EPR) [2, 4, 7]. Health care organizations are implementing EPRs incrementally (i.e. implementing components of the EPR over a period of years) [7, 8]. Even in technologically advanced countries where greater than 50% of EPR components have been implemented, the evidence suggests that a high proportion of organizations are implementing hybrid (i.e. electronic-paper) environments in their progression towards a full EPR. This has led some researchers to conclude that a full EPR may be difficult to implement, that paper may continue to exist, and that hybrid environments are best suited to the clinical setting [9]. As a result, some researchers have concluded the vision of a full EPR will be replaced by a hybrid environment. Similar conclusions have been made by researchers in other industries (e.g. banking) who have noted that it may not be possible to migrate to a fully paperless environment [10].

The discussion of hybrid environments in the research literature and their impact upon clinician information seeking has been limited. In studies, where there has been some attention drawn to aspects of information seeking (e.g. information needs) involving patient records, only full EPRs have been studied and not in the context of a clinical environment [11-12]. Furthermore, research has largely been descriptive, outlining the information needs of clinicians rather than studying the effects of the EPR upon information seeking. These studies have also failed to include other sources of information that are present in a typical clinical environment (e.g. textbooks, health professionals). Lastly, no attempts have been made to test for differences in clinical behaviour between paper and hybrid environments. Therefore, the purpose of this study was to investigate the effects of a hybrid environment upon health professional (i.e. clinician) information seeking (i.e. amount of information accessed, choice of key sources of information, types of information and use of information seeking tactics).

2. Methods

2.1. Sample and Materials

Thirty-five novice nurses participated in a laboratory based, experimental study (power = 92%; effect size = 93 for the primary outcome variable: information accessed). In this study, two representative, hospital-based environments were simulated. One was paper-based (i.e. all of the information in the environment was provided to subjects in paper form). The paper environment represented a nursing station where subjects are provided with paper-based information (i.e. a paper patient record, health professional reference texts, a communication book and a list of health professional colleagues that could be contacted by phone). The second simulated environment or the hybrid

environment (i.e. electronic-paper) provided subjects with access to a patient record, where part of the patient record was electronic and accessible via a workstation and the other part was paper-based and accessible via a paper chart. Patient demographic information, doctor's orders and documentation records were in the EPR and all other parts of the patient record and sources of information were paper-based (e.g. reference texts, a communication book and a list of health professional colleagues).

Two patient records were used in the study (i.e. A and B). The patient records were taken from a hospital during a time period when the hospital was fully paper-based. The patient records represented patients who were early in their hospital stay and had been admitted to a medical unit for one day (i.e. there was a high degree of uncertainty associated with their conditions to stimulate information seeking) [14]. Both the patient records represented older adults diagnosed with multiple chronic and acute illnesses. The patient records were selected by a Masters prepared, expert nurse who deemed them to be typical patients a novice nurse would encounter. Each record was in paper form and was also recreated in hybrid form for the hybrid environment. Both environments (i.e. the paper and hybrid) were evaluated for their ecological validity (by six experts: three health informatics and three clinician) and were deemed typical of the types of environments that are present in North American hospitals [8].

2.2. Study Design

A within group, laboratory, experimental study was conducted using the two simulated environments (i.e. paper and hybrid). Subjects were asked to perform the task of assessment and planning of a new patient's care involving the review of a new patient's record at the beginning of a nursing shift. Assessment and planning involving a review of a patient's chart stimulates nurse information seeking [13]. Subjects were presented with a paper and then a hybrid environment. Research has demonstrated the presence of long term carry over effects following the use of EPRs [12]. Therefore, subjects were block randomized to receiving either patient A's or B's information first (i.e. AB or BA) to prevent learning, order effects and effects associated with differential difficulty between the patients. Subjects were video and audio taped "thinking aloud" in both environments. Subjects participated in a cued recall session following the simulations. During the cued recall session, subjects were asked to clarify their verbalizations and/or their actions [14].

2.3. Data Analysis

Protocol analysis was used to code the data [14]. "Think aloud" data from the simulations was transcribed, and divided into segments according to natural pauses in speech. The data (i.e. audio and video) was then coded using Transana® according to information seeking concepts defined by Miller and Jablin's [3] and Johnson's [2] work (See Table 1). The unit of analysis for the data consisted of utterances, words, phrases, sentences, paragraphs, and video frames (i.e. the smallest unit of information that could be understood [11, 14] and could be represented by concepts defined by Miller and Jablin [3] and Johnson [2]). Ericsson and Simon refer to this coding approach as Model Based Coding [14]. A Kappa of 0.914 was calculated, indicating high in-coder agreement. The coded audio and video data from the simulations was then quantified [12]. The numbers of occurrences of each code were counted for each subject in both environments. Paired sample t-tests were conducted to test for differences between the

environments for the amount of information accessed, type of information, sources of information and use of information seeking tactics (i.e. aspects of information seeking). This method has been used by researchers who have quantified “think aloud” data to conduct inferential statistics [e.g. 11-12, 14].

Table 1. Coding Scheme for Theoretically Based Concepts

Code	Definition
Data Accessed	Item of data “attended to” or “heeded” by the subject [14].
Types of Information	
Referent	Information that “tells the worker what is required of him or her to function successfully in their work” [14, p.48].
Relational	Information that “tells the worker about the nature of their relationship with others” in the workplace [14, 1978].
Appraisal	Information that “tells the worker how well they are functioning” at work [14, p.48]
Information Seeking Tactics	
Active Tactics	Active information seeking tactics involve direct interaction with the information source or target (i.e. overt questions) [2]
Passive Tactics	Passive information seeking tactics involve indirect interaction with the information source (i.e. testing, indirect questioning, surveillance, observing, disguising conversations, third party questioning [2].
Sources of Information	
Paper	A paper source of information (e.g. a paper patient record, a text book) [7]
Electronic	An electronic source of information (e.g. an EPR) [7].

3. Results

There were a number of statistically significant differences in terms of subject information seeking between the two environments. In terms of the primary outcome variable (i.e. information accessed), subjects accessed less information in the hybrid than the paper environment. The eta squared statistic for information accessed was calculated as 0.3 which indicated a small to moderate effect size. Paired t-tests were conducted to test for differences between the paper and hybrid environment for subject use of referent, relational and appraisal information. There were no statistically significant differences between the two environments in terms of the use of these types of information. Subjects accessed more non-electronic than electronic sources of information in the hybrid environment. Tests for differences in subject use of active and passive information seeking tactics in both environments were also conducted. There were no statistically significant differences in active information seeking between the two environments. Alternatively, there was a statistically significant difference in passive information seeking. Passive information seeking tactics were used more frequently in the hybrid than the paper environment (See Table 2 for more details).

Table 2. Results

Aspect of Information Seeking	Environment	Mean	P value
Data Accessed	Paper	276.91	0.001***
	Hybrid	184.51	
Types of Information			
Referent	Paper	12.34	0.17
	Hybrid	9.57	
Relational	Paper	8.34	0.1
	Hybrid	6.37	
Appraisal	Paper	0.17	0.90
	Hybrid	0.20	
Information Seeking Tactics			
Active tactics	Paper	18.60	0.92
	Hybrid	18.46	
Passive Tactics	Paper	1.09	0.04***
	Hybrid	1.11	
Sources of Information			
Electronic versus Non-electronic	Electronic	7.58	0.001***
	Non-Electronic	4.09	

*** indicates statistical significance

Cued recall data suggested 25 subjects' (i.e. 74%) experienced increased cognitive load or difficulty integrating information from differing sources of information (i.e. the paper patient record, the EPR, textbooks, communication book and/or a list of health professionals that could be contacted) in the hybrid condition. Subjects described the experience of obtaining information from information sources in the hybrid environment as "frustrating" and "confusing" when going back and forth between the EPR and the paper record and other paper-based information resources; for example:

I would like to have all of my information in one place then I am not transferring back and forth. It was frustrating to have to check on the computer. Check everywhere on the computer. Oh it is not there. Then go back to the chart and oh its not there either and then try it again [29, Cued Recall]

In summary subjects accessed more information in the paper environment. Subjects also accessed more non-electronic sources of information than electronic sources in the hybrid environment. Lastly, passive information seeking tactics were used more frequently in the hybrid environment than the paper environment by subjects. 74% of subjects experienced cognitive load in the hybrid environment.

4. Conclusion

The purpose of this study was to determine if hybrid environments can influence aspects of clinician information seeking. In order to determine the effects of a hybrid

environment upon information seeking, two typical hospital environments were simulated. Study findings revealed the hybrid environment altered the amount of information accessed by information seekers. Subjects accessed less information in the hybrid than the paper environment. Cued recall data suggested subjects may have experienced increased cognitive load arising from the need to consult more sources of information in the hybrid environment. This may have placed excess demands upon subjects' cognitive processes such as memory causing them to access less information. The implications of this work are significant. Health informaticians will need to take into consideration the way they implement and train health professionals to use hybrid environments as such environments may lead to increased cognitive load and the use of less information involving patient care. Future research needs to explore how the order of implementation of the components of the EPR impact upon cognitive load and clinician ability to seek information in hybrid health care environments.

References

- [1] L. M. Currie, M. Graham, M. Allen, S. Bakken, V. Patel, J. J. Cimino, Clinical information needs in context: An observational study of clinicians while using a clinical information system. In M. A. Musen, C. P. Friedman, J. M. Teich (Eds.), *AMIA 2003 Symposium: Foundations of Informatics: Vol. 27*. (pp. 190-194). American Medical Informatics Association, Washington, D. C., 2003.
- [2] J. D. Johnson, *Information seeking: An organizational dilemma*, Quorum, Westport, Connecticut, 1996.
- [3] V. D. Miller, F. M. Jablin, Information seeking during organizational entry: Influences, tactics and a model of process, *The Academy of Management Review* **16** (1991), 92-120.
- [4] Accenture, *Managers say the majority of information obtained for their work is useless: Accenture survey finds*, Retrieved March 17, 2007. http://newsroom.accenture.com/article_print.cfm, 2007.
- [5] J. G. March, H. A. Simon, *Organizations*, John Wiley and Sons, New York, 1958.
- [6] D. G. Covei, G. C. Ulman, P. R. Manning, Information needs in office practice: Are they being met? *Annals of Internal Medicine* **103**(4) (1985), 596-599.
- [7] E. H. Shortliffe, J. J. Cimino, *Biomedical informatics: Computer applications in health care and biomedicine*, Springer Verlag, New York, 2006.
- [8] Branham Group, *eHealth in Canada: Crossing the chasm*. Branham Group, Ottawa, 2007.
- [9] R. Pizzi, *Hybrid medical records an option for some hospitals*. Healthcare IT News, New Gloucester, Maryland, **6** (2007),1.
- [10] A. J. Sellen, H. R. Harper, *The myth of the paperless office*. MIT Press, Cambridge, Massachusetts, 2003.
- [11] A. W. Kushniruk, D. Kaufmann, V. L. Patel, Y. Levesque, P. Lottin, Assessment of a computerized patient record system: A cognitive approach to evaluating medical technology, *MD Computing*, **13**(5) (1996), 406-412.
- [12] V. L. Patel, A. W. Kushniruk, S. Yang, J. F. Yale, Impact of a computer-based patient record system on data collection, knowledge organization, and reasoning, *JAMIA*, **7**(6) (2000), 569-585.
- [13] S. Salentera, E. Eriksson, T. Junnola, E. K. Salminen, S. Lauri, Clinical judgment and information seeking by nurses and physicians working with cancer patients. *Psycho-onology*, **12** (2003), 280-290.
- [14] K. A. Ericsson, H. A. Simon, *Protocol analysis: Verbal reports as data* (2nd ed.). MIT Press, Cambridge, Massachusetts, 1993.
- [15] L. M. Hanser, P. M. Muchinsky, Work as an information environment. *Organizational Behaviour*, **21** (1978), 47-60.

Application of Business Process Management to drive the deployment of a speech recognition system in a healthcare organization

María José GONZÁLEZ SÁNCHEZ^{a,1}, José Manuel FRAMIÑÁN TORRES^b, Carlos Luis PARRA CALDERÓN[‡], Juan Antonio DEL RÍO ORTEGA^a, Eduardo VIGIL MARTÍN^a, Jaime NIETO CERVERA^a

^a *'Virgen del Rocío' University Hospitals. Seville. Spain.*

^b *School of Engineering. University of Seville. Seville. Spain*

Abstract. We present a methodology based on Business Process Management to guide the development of a speech recognition system in a hospital in Spain. The methodology eases the deployment of the system by 1) involving the clinical staff in the process, 2) providing the IT professionals with a description of the process and its requirements, 3) assessing advantages and disadvantages of the speech recognition system, as well as its impact in the organisation, and 4) help reorganising the healthcare process before implementing the new technology in order to identify how it can better contribute to the overall objective of the organisation.

Keywords. Human Interfaces, Speech Recognition, User interfaces, User-computer interface.

1. Introduction

It is well-known that healthcare practice produce large amounts of documentation. Researches conducted by the American Medical Industry Transcription Association in 2006 and by the Giga Information Group in 2004 estimate that around 12.000 millions dollars a year are spent in transcription of medical report only in the USA [1;2]. The process of dictating and transcribing these reports and delivering them through the hospital and to patients is a slow and troublesome task. Nowadays, the most generalized way to fill these reports is the following: the doctor records an audio tape that it is transcribed by administrative staff. The doctor, who revises and signs it, then examines the output of this transcription. In this way, it can take several days to

¹ Corresponding address: María José González, Grupo de Investigación y Desarrollo Tecnológico, Hospitales Universitarios "Virgen del Rocío". Avenida Manuel Siurot, s/n. Edificio Documentación Clínica Avanzada. CP: 41013. Sevilla. España; E-mail: mjose.gonzalez.exts@juntadeandalucia.es

complete the medical reports, which may delay patients' treatment. This problem is becoming more critical due to several factors, including the following:

- The aging of the population, since a bigger amount of reports, tests and treatments are required[3] , and
 - the increasing need of having electronic documentation that can be shared [3].
- Medical practitioners need to get the patient information in an immediate and exact way to ensure optimal results [2].

To solve this problem, speech recognition systems have been proposed as a new technology able to automate the capture of the information, thus allowing its distribution in a nearly immediate way. Speech recognition systems can eliminate problems derived from transcriptions, the verbal communication, or communication through manuscript notes between professionals [1-12].

The implementation of speech recognition systems may imply an important reorganization of the resources in the hospital, as it serves to reduce or eliminate subcontracting external services and to release the internal administrative staff of the tedious tasks of typing [2; 11]. To carry out this implementation, a description on the way that doctors conduct their work and their technical needs is required in order to adapt the technology to their environment. Every Medical Unit may work in a different way and every doctor acts according to his/her experience, knowledge and habits, therefore the elicitation of these processes is very difficult, particularly for the non-medical staff in charge of the deployment of the speech recognition system.

To overcome these difficulties, in this paper we propose the application of Business Process Management (BPM) techniques. BPM aims to improving the efficiency of a system by means of modelling, automating, to integrating, and optimizing business processes in a continuous manner. There are several goals behind BPM, including creating new and better processes, understanding what is done well or badly, process automation, and creating and supporting value chains [13; 14; 15].

The first step in BPM is modelling the processes, i.e. developing a formal description of a system and the activities carried out (the so-called 'as-is' process). By modelling the process, the existing interrelationships among different activities, events, stakeholders and resources are depicted so existing problems can be identified and possible improvements can be suggested. Next, a model incorporating these improvements (the so-called 'to-be' model) is built. The construction of the 'to-be' model allows analysing its potential advantages with respect to the 'as-is' model by means of discussions with stakeholders in the process, discrete-event simulation, etc. so an 'a priori' evaluation of the model can be conducted [16].

In industry, BPM is regarded as a very efficient tool, ensuring the success in a high percentage of the companies and organizations where it has been used [17]. BPM capabilities will enable care delivery organizations to improve efficiency and reduce medical errors [16]. Although not very known in the health care domain prior to 2001 [18], nowadays there are well-documented applications, such as e.g. the choice by the Veterans Health Administration (VHA) of a BPM platform as its technology solution for Case Management of VHA Benefits Eligibility.

This paper describes the experience of the application of BPM within the development of the VOZENEC (*VOZ EN Estación Clínica* – Voice in a Clinical Station) project. This project consists on the deployment and integration of a speech recognition system to convert speech to text within the Hospitales Universitarios 'Virgen del Rocío', the largest hospital in Andalusia, Spain. Several pilot Units were

selected for the project, i.e. Radiology and Pathological Anatomy, Respiratory Diseases Unit, Plastic and Reconstructive Surgery Unit, and Endocrinology and Nutrition Unit.

The rest of the paper is organised as follows: next we summarize the methodology and describe the output of our experience, focusing on the benefits gained by the application of the BPM methodology. We conclude presenting areas for future work.

2. Methods

The methodology adopted for process modelling is taken from ARIS (ARchitecture of Integrated Systems). The choice of the ARIS methodology has been based on a previous analysis of the existing methodologies for business process modelling, being this one the methodology that more closely feels the needs presented in the healthcare processes. Among them, it is worth to note the fact that it allows the description of both the processes that specify each activity and the underlying processes structure, the flow of objects and its relations [21].

The core of the methodology consists on the following phases:

1. Process definition: In this phase, the healthcare processes to be studied are established together with a definition of its limits, inputs and outputs, Units involved in the execution of the processes, selection of the staff in charge of assisting during the modelling process and in the validation of the models.
2. Building static 'as-is' models: The output of the phase is a detailed graphical representation of the process in terms of a formal modelling language named Event-driven Process Chains or EPCs. In our case, it refers to the processes before the implementation of the speech recognition. The 'as-is' models were obtained through an iterative process of design and verification until their validation by the staff in the different Units.
3. Building dynamic 'as-is' models: By extracting data from the Hospital Information systems data as well as data provided by the staff in the different Units, discrete-event simulation models of the 'as-is' processes have been built. The results of the simulations have been analysed and validated by the staff.
4. Building 'to-be' models: In our case, it consists on the expected final state after the implementation of the speech recognition system. These models can be simulated so the impact of the technology in the process can be evaluated, thus allowing detection of potential problems, estimation of benefits and disadvantages, identification of new activities that should be improved in the future, etc.

3. Results

Several advantages have been identified by driving the deployment of the speech recognition system by the BPM methodology discussed in the previous section. We group them into three types:

- Development of the to-be model. By involving the key staff within the units in the modelling processes (see steps 2 and 3 in the previous section), they have now a clear picture of all relevant processes in the unit, not only of those under their direct responsibility. As a result, they gained a bigger knowledge of the clinical practice in their units and they were able to detect points for

future improvements they were not aware previously. This is a widely-discussed characteristic of BPM of particular interest in the healthcare sector, as the strong specialization of the clinical staff has led to the fragmentation of the healthcare processes into small areas of responsibility where (local) improvements may not be reflected in the (global) process. As a result, some practices have changed in view of the implementation of the speech recognition system. For instance, in the Pathological Anatomy Service, the macroscopic report used to be sent to the pathologist's consulting room while waiting for the laboratory's results. In the new process, the report is sent to the laboratory together with the specimens. When the results are obtained from the laboratory, then both the report and the results are sent to the corresponding pathologist for diagnosis. With this organizational change, administrative (non-value added) time has been saved with respect to the previous situation. Also, by means of the study and modeling of Macroscopic processes in the Service of Pathological Anatomy it is chosen to use digital tape recorders instead of using microphones connected to PCs, as the hands of the staff were occupied for most of the time.

- A priori evaluation of the new technology. The BPM methodology employed here has served to make an a priori assessment of the technology to be implemented regarding both time and cost. By simulating the processes and analyzing the output, important time reductions regarding document generation have been anticipated. For instance, the time reduction in document generation was estimated to be about 89% with respect to the previous situation in the Endocrinology and Nutrition Unit. These estimations proved to be fairly accurate when the technology was implemented, as well as the costs reductions estimated by the simulation of the processes.
- Finally, process modeling has eased the deployment of the speech recognition system in the different Units, as the documentation and analysis of the process gained through process modeling have made the IT professionals in charge of the implementation of the speech recognition system to exactly know the requirements and daily practices of the medical practitioners, thus reducing both the time for deployment and its degree of success. Also, a number of documents and templates have been obtained, being the most relevant a reference model of the BPM methodology so its different phases are described, as well as the participants implied in the different stages. Finally, a glossary of terms has been developed. Both the reference model and the glossary have been validated and tested through practice, so they constitute useful tools to help deploying more efficiently the speech recognition system in the rest of the units in the hospital

4. Conclusions

A speech recognition system allows reducing the time devoted to transcribing the reports, and thus the time to complete the whole healthcare process from the patient's viewpoint may be accelerated. It also serves to automate and to centralize the process of documentary generation [1; 2; 4-8; 11; 12; 16; 17], which in turns facilitate the use of Electronic Clinical Records. On the other hand, a speech recognition system involves the introduction of a new technology in an organization, which it turn may

imply an important reorganization of the resources in the hospital, as it serves to reduce or eliminate subcontracting external services and to release the internal administrative staff of the tedious tasks of typing [2; 11]. In order to efficiently cope with the organisational change required by the implementation of such a system, we suggest a methodology based on BPM. The proposed methodology is general in the sense that it can be employed for evaluating and driving the deployment of any kind of Information Technology. The results show that the application of this methodology has been useful and beneficial, and we have been able to fulfil the aforementioned needs. In addition, we must highlight the satisfaction of the clinical staff in the Units, who have gained a comprehensive knowledge of their units via their involvement in helping to model the processes.

The experience and positive results obtained in this project have revealed a number of issues worth of further research. One of the most prominent is the possibility of applying this technology to facilitate the procedure of codification of the Electronic Patient Records, i.e. to develop a system that could codify automatically the clinical information that is generated from the interface of the speech of Clinical Station. The BPM proposed methodology could be used then to identify the requirements posed by the different actors in the process.

5. References

- [1]. Bajaj S. Expanding the reach of speech recognition. Extending speech technology benefits to transcription workflow can boost productivity. *Health Manag Technol* 2004; 25(3):16-22.
- [2]. Gestión de Documentación médica. 2006. www.speechrecognition.philips.com.
- [3]. Nick van Terheyden. Is Speech-Recognition the Holy Grail? *Health Management Technology*; 2005; 26,2:42-44.
- [4]. Robin Blair. Bye-bye, Foot Pedal - Hello, Efficiency. *Health Manag. Technol.* 40-42. 2005.
- [5]. Joseph Goedert. Is the speech recognition market finally ready for growth? 10,2, 12-13. 2002. *Health Data Management*.
- [6]. Suzanna Hoppszallern. Making Stat Faster. *Hospitals & Health Networks*; 5,1, 76. 2006.
- [7]. Christopher Spring. Productivity Gains of Speech-Recognition Technology. *Health Manag Technol* 2003; 24,1:54-55.
- [8]. Software de reconocimiento de voz al servicio de la sanidad en Valencia. <http://www.lavanguardia.es/gen/20060828/51281148928/noticias/software-de-reconocimiento-de-voz-al-servicio-de-la-sanidad-en-valencia-generalitat-rafael-blasco.html>.
- [9]. Mark S Lucas. Speaking of Efficiency. *Health Management Technology*; 2001; 22,7:42-44.
- [10]. Speech Magic. El reconocimiento del habla en el que confían los profesionales. 2005. www.philips.com/speechrecognition.
- [11]. Tecnología de Reconocimiento de Voz. Agfa HealthCare incorpora una tecnología de reconocimiento de voz de eficacia demostrada. 2004. www.agfa.com/healthcare.
- [12]. Joanne Galimi CEBJY. Underlying Information Technologies in Healthcare in 2004. ID Number: RPT-0404-0034. 2004. Gartner.
- [13]. Business Process Management.. 2006. http://searchcio.techtarget.com/sDefinition/0,,sid182_gci1088464,00.html
- [14]. BPM Basics.2008. <http://www.bpmbasics.com/literature/whatisbpm.html>
- [15]. New Direction in BPM Tools and Technology. Bruce Silver February 2007. <https://www.bpm institute.org/uploads/media/Silver-4-10-07.pdf>
- [16]. Barry R. Hieb, M.D. BPM/Workflow Boosts the Value of Computer-Based Patient Record System 2005. Gartner
- [17]. Bustos C. La Reingeniería: Herramienta controversia. *Vision Gerencial Año 4* 4[1]. 2005.

- [18]. Osorio Acosta J PAE. Reingeniería de procesos en los hospitales públicos: ¿Reinventando la rueda? *Revista Española de Salud Pública* 75[3], 193-206. 2001.
- [19]. University College London Hospitals Foundation Trust Deploying Lombardi Business Process Management (BPM) Suite for...2007. <http://www.allbusiness.com/health-care/health-care-facilities-hospitals/5268583-1.html>
- [20]. Veteran's Health Administration Selects HandySoft's BizFlow as Backbone for Benefits Eligibility Case Management System. 2005
<http://www.biohealthmatics.com/News/PressReleases/2005/03/08/00000000922.aspx>
- [21]. Carlos Parra Calderón, José Manuel Framiñán, Margarita Montes Worboys, Paz Pérez González. *Guía de estilo de Modelado de Procesos de Salud*. 2005.
- [22]. Reconocimiento del habla: Acelerando la creación de informes médicos. 2006.
<http://www.speechrecognition.philips.com/index.asp?id=747>.
- [23]. Testimonios de facultativos que usan reconocimiento de voz y servicios dónde se ha implementado. 2006. <http://www.speechrecognition.philips.com>.

Affective Computing and Medical Informatics: State Of The Art in Emotion-Aware Medical Applications

Andrej LUNESKI^{1,2}, Panagiotis D. BAMIDIS², Madga Hitoglou-Antoniadou³

¹*South-East European Research Centre (SEERC), Thessaloniki, Greece*

²*Lab of Medical Inform., Medical School, Aristotle Univ. of Thessaloniki, Greece*

³*Lab of Speech and Communication Disorders, 1st ENT Clinic, Medical School, Aristotle Univ. of Thessaloniki, Greece*

Abstract. The area of affective computing has received significant attention by the research community over the last few years. In this paper we review the underlying principles in the field, in an effort to draw threads for possible future development within medical informatics. The approach is lead by considering the three main affective channels, namely, visual, audio/speech, and physiological in relation to e-health, emotional intelligence and e-learning. A discussion on the importance of past and present applications together with a prediction on future literature output is also provided.

Keywords. Affective Computing, HCI, Emotion, Health

Introduction

Emotions play a significant role in the expression of human intelligence [1]. The arguments in support of the significance of emotions introduced a new area of “emotional intelligence”, defined as “the capacity to understand emotional information and to reason with emotions” [1]. Emotional intelligence is covered by the broader area of affective computing (AC), which represents “computing that relates to, arises from, or deliberately influences emotions” [2]. It introduces new domains and applications that can provide beneficial advancement of current technologies in healthcare and medical informatics, among others.

In the everyday interaction with computers, people exhibit emotions that in certain manners might influence their health. These emotions are communicated through three channels: audio (speech), face and body gestures (visual) and internal physiological changes (blood pressure, heart beat rate, respiration, skin sweating etc.). Identification of the human emotional displays by the computer requires for it to monitor the user, and based on certain parameters, classify/recognize the current emotional state. Each of the three channels of emotional interaction by the user and the computer has its own characteristics and applications in the healthcare domain.

This paper provides a summary of the research and the state of the art in affective computing and its impact to human health. The following section describes the relation

between emotions and health, as the chief reason for the recent focus into the area of affective computing by the medical informatics society. Furthermore, there is an explanation of how each of the three channels of affective computing is applied in the current medical technologies and the resulting benefits. The fourth section examines the significance of affective computing for the emerging and advanced technologies. Finally, there is a general discussion of the present status of affective computing in the healthcare domain and future directions for its advancement.

1. Emotions and Health

It has been a known fact that emotions have an un-bound relation with human health. The impact emotions have on it has been reported for centuries, even in religious writings: “The joyfulness of man prolongeth his days”, (Bible, Ecclesiasticus.30:22) or by ancient Greek philosophers and scientists, such as Socrates and Hippocrates, who considered emotion as a determinant of human health and diseases [3]. The advancement of technology has led to major changes in a human’s life, a more dynamic lifestyle, with more unpleasant stressful situations, frustration, irritation, depression and other emotions with strong negative impact to the human health. It has been noted in [4] that helplessness and negative emotions such as stress or depression weaken the human immune system. Stress causes faster and harder heart pumping, which can lead to the amount of blood reaching heart through the coronary arteries not to be enough to support the faster work of the heart, or a condition known as myocardial ischemia [6]. Defence mechanisms are essential for an individual’s ability to cope with negative emotions such as depression and anxiety [7]. Stressful emotions can have an enormous impact on an individual’s mental health and therefore the strong adaptiveness of defence mechanisms can protect from the negative health effect.

Positive emotions highly contribute to physical and mental health. Laughter generates positive emotions, which in turn help in the improvement of the functioning of the immune system [8]. Humour and laughter act as a primary defence in stressful situations and research has shown that individuals that are able to preserve their positive mood with humour have stronger immune systems [9]. Positive emotions can also act as a medicine for cardiovascular diseases. Middleton & Byrd [10] state that elderly patients suffering from cardiovascular diseases, that were happier than others, had a lower number of readmissions to hospitals. Mental and physical health are influenced by self-esteem and self-efficacy [11], while optimism can speed up recovery from heart surgery and breast cancer [12], enhance greater social integration by increasing the positive “self-view” and happiness, prolong life and assist in coping with stressful situations [13].

The awareness of the significance of emotions and their impact on human health has boosted the motivation for improvement of the research in this field. Together with the advancement of affective technologies in the last two decades, a new paradigm for health has emerged; a paradigm that we feel may enforce the field of medical informatics to seriously take into consideration AC technologies and the evident benefit of their applications.

2. Affective Speech

Emotions in speech can be expressed semantically and through speech prosody. While semantics (what has been said) is a more obvious way of identifying one's emotional state, prosody can provide more detailed information. The term prosody combines nonsemantic cues in spoken language, such as: fundamental frequency (pitch), rhythm, loudness, intonation, formant structure of speech sounds etc.

Research into emotionally expressive speech is very significant for people with Asperger syndrome. Despite normal early language development, individuals with AS are often characterized by abnormal prosody and impaired semantics and pragmatics as well as poor social skills and emotional behaviour [14].

Affective prosody has been employed as a feature for studying the Williams Syndrome (WS), as well. The effects of WS are low non-verbal IQ, impairments in planning and problem solving, uneven cognitive profile linguistic disabilities etc. [15]. Reilly et al. have examined the affective vocal prosody on children with WS through a story telling experiment. It was noticed that the same level of expressiveness was used by the adolescent speakers with WS, regardless of the number of repeats of the story telling or the number of the audience.

An extensive number of studies, concerned with the Parkinson Disease (PD) have reported that patients with PD fail to recognize the emotional content in prosody [16]. Therefore, experiments have been performed where affective speech has been used to further investigate this impairment [17].

3. Affective Psychophysiology

Most of the disturbance caused by a certain emotion influences the internal physiology of the human body (increase of the heart-beat rate, palm sweating, increased respiration rate etc.). Monitoring of these types of emotional expressions requires employment of specific sensors such as Electroencephalogram (EEG), Electromyogram (EMG), Electrocardiogram (ECG), sensors measuring Electrodermal Activity (EDA) etc. The usage of physiological signals in identifying a user's emotional state has become popular in recent years due to the advanced development and availability of unobtrusive sensors that can provide constant and reliable monitoring of a user's internal emotional reaction [3]. Physiological sensors have been successfully integrated into clothing and jewelry; skin conductivity sensor in shoes, blood volume pressure sensor in earrings, respiration sensor in a sports bra and numerous others [18]. These sensors have enabled monitoring of users under various every-day conditions. Haley and Picard have conducted experiments for monitoring drivers' physiologic reactions during real-world driving situations under stress in normal every day surroundings [19]. Bamidis et al. have proposed a Multi-channel framework for experimenting with physiological sensing of human emotions [20].

Monitoring of patients' emotional state using physiological signals has been a primary feature of the existent and emerging applications in tele-home healthcare and ambient assistive living (discussed further in the remainder of this paper).

4. Affective Facial Expressions and Gestures

Facial expressions are mostly used by humans in identifying certain categories of emotions due to the distinct facial differences for the “universal” human emotions such as happiness, anger, sadness, surprise, fear and disgust [21]. The computer needs to be able to recognize human facial emotions as well as express them in the most natural way possible. The latter can be accomplished through artificial robots or through virtual human characters (avatars) that can mimic the human facial expressions. Some of the most characteristic applications are concerned with the ability of the computer to respond to user frustration by expressing empathetic emotions through facial expression of an avatar [22]. Apart from the above-mentioned affective speech, emotionally-expressive avatars have also been used in therapies and learning experiments for autistic people [23].

5. Affective Computing in the Medical Informatics Domain

Once equipped with such sensors, the computer is capable of identifying/recognizing emotions or emotion categories. The latter sets the base for the general vision of the benefits of AC in medicine – What can computers do and how can they be of help, once they have detected the correct emotional state of the user? Researchers have only recently started to think of possible answers to such questions and there is already a significant progress in several medical areas.

Tele-home health care (THHC) is one of the areas that have received great attention by the AC community. Internet-based communication technologies have enabled patient monitoring without the need of physical presence by the caregiver. Current systems facilitate collection of vital sign data remotely, such as ECG, blood pressure, oxygen saturation, heart and breath sounds, verification of compliance with medicine regimes, assessment of mental or emotional status and more [25]. Communication between the caretaker and care recipient through emotional channels in such environments has shown to be of vital importance to the patient [26]. Existing systems employ multi-modal interface including avatars (able to express emotions), to remind the patient for a medication, show empathy to the user when certain negative emotion is detected etc [27].

AC, recently, has been included in the research in the area of Ambient Intelligence (AmI), also referred to as Pervasive or Ubiquitous Computing. Picard considered AmI as one of the most important emerging technologies that will be interrelated with affective computing in the future [3]. The significance of AmI, together with AC, has been reported in the final report by the IST Advisory Group (ISTAG) regarding “Scenarios for Ambient Intelligence in 2010” [28]. Moreover, Riva examines the AmI implications on the future of health technologies, stating that AmI has enormous affect on technology, ergonomics, project management, human factors and organizational changes in the structure of the relevant health service [29]. Recent research attempts in AC have been concerned with the learning and education fields. The basic idea behind this initiative is the influence of emotions and mood on the learning performance and decision-making process. Picard and the Affective Group are working on Affective Learning Companion¹ project that uses software-based interactive application that will

¹ <http://affect.media.mit.edu/projectpages/lc/>

recognize the affective and cognitive state of the learner and respond in an appropriate manner (e.g., can adjust the pace, difficulty, complexity). The existent affective technologies such as virtual agents that can visually express emotions, affective speech synthesis and recognition etc, enable the successful application of AC in e-learning.

6. Discussion

Research has shown that the medical community has started realizing the crucial role emotions play in the preservation of human's mental and physical health. As reported in this paper, there is an enormous increase in studying the relation between emotions and health (medicine) in recent years. According to [4], the interest in emotions and health has been highly increased since 1991, and especially in the last 10 years. We have made an effort in adding statistics for the 2001-2007 period and estimation for the overall 2006-2010 period. In 1991, there were 2000 searches on the "emotion and health" topic, while in 2006 we have nearly 6000. The estimation is constructed linearly by taking into consideration the growth of the previous statistical data, and the number of around 3000 publications from 01/01/2006 until 10/2007. However, it has to be noted that this estimation is rather simple and cannot be considered as completely accurate; it merely illustrates the growing awareness and interest on emotions by the medical community. Additionally, even though the quantity of published papers cannot by itself be an accurate indicator of the importance of this emerging area, the boost in the published research works might indirectly indicate that researchers have decided to pay more attention into this field and, therefore, it definitely shows potential for future applications.

Since its emergence in the late 90's, AC has provided the medical community with technologies that help with better understanding of emotions, identifying their impact on health, and offering new techniques for diagnosis, therapy, and treatment of emotionally-influenced diseases. This paper has provided insight into the most representative affective technologies and their current practice into medical informatics. Moreover, we have mentioned the tele-home healthcare, ambient intelligence and e-learning as areas where the potential of AC has already been realized and initial applications have emerged. Ambient assistive leaving, which incorporates both tele-home and ambient intelligent technologies has been included in the ICT FP7 Work Programme under Challenge 7 – ICT for Independent Living and Inclusion². However, utilization of AC in medical practice is only in its infant phase. Many domains are yet to be explored. For, instance medical education, through the area of e-learning can benefit from AC. Affective speech synthesis and multi-modal emotion expression through virtual characters have started to become popular in virtual community applications for elderly people or even children with certain impairments. The number and variety of AC applications in the medical domain is dependent on the development pace of AC technologies. Therefore, with advances in each of the three sub-areas of AC (speech, face and gestures expression, physiology), we can expect an enormous increase in the interest for emotionally-intelligent applications in the medical informatics domain.

² <http://cordis.europa.eu/fp7/ict/>

7. References

- [1] P. Salovey and J.D. Mayer, "Emotional Intelligence", *Imagination, Cognition and Personality*, vol. 9, pp. 185-221, 1990.
- [2] R. W. Picard, *Affective Computing*, MIT Press, Cambridge, MA, Sept. 19th, 1997
- [3] J.T. Cacioppo, "Introduction: Emotion and Health", in *Handbook of Affective Stress*, 1st ed., editors: Richard J. Davidson, Klaus R. Scherer, Hill H. Goldsmith, Oxford University Press, 2003.
- [4] R. Glaser, D. K. Pearl, J. K. Kiecolt-Glaser W. B. & Malarkey, "Plasma cortisol levels and reactivation of latent Epstein-Barr virus in response to examination stress", *Psychoneuroendocrinology*, vol. 19, pp. 765-772, 1994.
- [6] M. M. Burg, "Stress, behavior, and heart disease", in: *Heart Book*, B. L. Zaret, M Moser, L. S. Cohen eds., Yale University School of Medicine, 1992.
- [7] D. L. Paulhaus, B. Fridhandler, S. Hayes, "Psychological defense: Contemporary theory and research", in *Handbook of personality psychology*, R. Hogan, J. Johnson, S. Briggs eds, Academic Press, San Diego, 1997.
- [8] J. Bachorowski & M. J. Owren, "Not all laughs are alike: Voiced but not unvoiced laughter readily elicits positive affect", *Psychological Science*, vol. 12, pp. 252-257, 2001.
- [9] A. A. Stone et al. "Evidence that secretory IgA antibody is associated with daily mood", *Journal of Personality and Social Psychology*, vol. 52, pp. 988-993, 1987.
- [10] R. A. Middleton, E. K. Byrd, "Psychosocial factors and hospital readmission status of older persons with cardiovascular disease", *Journal of Applied Rehabilitation Counseling*, vol. 27, pp. 3-10, 1996.
- [11] A. O'Leary, "Self-efficacy and health: Behavioral and stress-physiological mediation", *Cognitive Therapy & Research*, vol. 16, pp. 229-245, 1992.
- [12] M. F. Scheier et al., "Dispositional optimism and recovery from coronary artery bypass surgery: The beneficial effects on physical and psychological well-being", *Journal of Personality and Social Psychology*, vol. 57, pp. 1024-1040, 1989.
- [13] D. J. Wiebe & P.G. Williams, "Hardiness and health: A social psychophysiological perspective on stress and adaptation", *Journal of Social and Clinical Psychology*, vol. 11, pp. 238-262, 1992.
- [14] E. Hill, S. Berthoz & U. Frith, "Brief report: Cognitive processing of emotions in individuals with autistic spectrum disorder and their relatives", *Journal of Autism and Developmental Disorders*, vol. 34, no. 2, pp. 229-235, 2004.
- [16] J. Reilly, E. S. Klima & U. Bellugi, "Once more with feeling: affect and language in atypical populations", *Development and Psychopathology*, vol. 2, pp. 367-391, 1990.
- [17] C. Breitenstein, I. Daum, & H. Ackermann, "Emotional processing following cortical and subcortical brain damage: contribution of the fronto-striatal circuitry", *Behavioural Neurology*, vol. 11, pp. 29-42, 1996.
- [18] C. Dara, L. Monetta, M. D. Pell, Vocal emotion processing in Parkinson's disease: reduced sensitivity to negative emotions, *Brain Research*, 2007, doi: 10.1016/j.brainres.2007.10.034
- [19] R. W. Picard, "Affective Medicine: Technology with Emotional Intelligence", in *Future of Health Technology*, R. G. Bushko eds., OIS Press, 2001.
- [20] J. Healey & R. W. Picard, "Detecting Stress During Real-World Driving Tasks Using Physiological Sensors", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 6, pp. 156-166, 2005.
- [21] P. D. Bamidis, A. Luneski, A. Vivas, C. Papadelis, N. Maglaveras, C. Pappas, "Multi-channel physiological sensing of human emotion: insights into Emotion-Aware Computing using Affective Protocols, Avatars and Emotion Specifications", *Proceedings of Medinfo 2007*, Brisbane, Australia, pp. 1068-72..
- [22] P. Ekman, *Emotion in the Human Face*, Cambridge University Press, New York, NY, 2nd edition, 1982.
- [23] J. Klein, Y. Moon and R. W. Picard, "This computer responds to user frustration: theory, design and results", *Interacting with Computers*, vol. 14(2), pp. 119-140, 2002
- [24] D. J. Moore, Y. Cheng, P. McGrath, N.J. Powell, 'Collaborative virtual environment technology for people with autism'. *Focus on Autism and Other Developmental Disorders*, vol. 20, pp. 231-243
- [25] A. Allen, et al., "Home health visits using a cable television network: user satisfaction", *Journal of Telemedicine and Telecare*, vol. 2, pp. 92-94, 1996.
- [26] C. Lisetti et al. "Developing multimodal intelligent affective interfaces for tele-home health care", *International Journal of Human-Computer Studies*, vol. 59, pp. 245-255, 2003.
- [27] H. Prendinger & M. Ishizuka, "What Affective Computing and Life-like Character Technology Can Do for Tele-Home Health Care", *Workshop on HCI and Homecare: Connecting Families and Clinicians*, in *conj. with CHI-04*, Vienna, Austria, 2004.
- [28] K. Ducatel, M. Bogdanowicz, F. Scapolo, J. Leijten & J-C. Burgelman, "Scenarios For Ambient Intelligence in 2010", *ISTAG Final Report*, February 2001.
- [29] C. Riva, "Ambient Intelligence in Health Care", *CyberPsychology & Behavior*, vol. 6, No. 3, 2003.

Using Medline Queries to Generate Image Retrieval Tasks for Benchmarking

Henning MÜLLER^{ac}, Jayashree KALPATHY-CRAMER^b, William HERSH^b,
Antoine GEISSBUHLER^a

^a*Medical Informatics Service, University & Hospitals of Geneva, Geneva, Switzerland*

^b*Oregon Health and Science University, Portland, OR, USA*

^c*Business Information Systems, University Of Applied Sciences, Sierre, Switzerland*

Abstract. Medical visual information retrieval has been a very active research area over the past ten years as an increasing amount of images is produced digitally and made available in the electronic patient record. Tools are required to give access to the images and exploit the information inherently stored in medical cases including images. To compare image retrieval techniques of research prototypes based on the same data and tasks, ImageCLEF was started in 2003 and a medical task was added in 2004. Since then, every year a database was distributed, tasks developed, and systems compared based on realistic search tasks and large databases. For the year 2007 a set of almost 68'000 images was distributed among 38 research groups registered for the medical retrieval task. Realistic query topics were developed based on a log file of Medline. This log file contains the queries performed on Pubmed during 24 hours. Most queries could not be used as search topics directly as they do not contain image-related themes, but a few thousand do. Other types of queries had to be filtered out as well, as many stated information needs are very vague; for evaluation on the other hand clear and focused topics are necessary to obtain a limited number of relevant documents and limit ambiguity in the evaluation process. In the end, 30 queries were developed and 13 research groups submitted a total of 149 runs using a large variety of techniques, from textual to purely visual retrieval and multi-modal approaches.

Keywords. Content-based image retrieval, classification, medical image retrieval, benchmarking, topic development

Introduction

The availability of large amounts of medical images in the electronic patient record and their use by clinicians who are often no specialists in radiology creates a need to develop new tools. Content-based image retrieval (CBIR) is a technique that aims at retrieving images based on their visual content instead of textual metadata [1]. In the non-medical domains this technique is frequently used to access and browse in large visual information repositories. It is used to complement text-based search particularly when little metadata is available. In the medical domain image retrieval has been discussed as important for a fairly long time [2,4]. It was used for image classification [3] as well as for retrieval with a large variety of visual features such as shapes, colors,

and textures [4,5,6]. First clinical tests also showed that diagnosis performance can be increased, particularly for non-specialists by making available additional data [7].

As medical image retrieval is in the process to transform from research prototypes towards real applications it is important to separate good techniques from techniques performing poorly. For this goal of technology assessment, benchmarks have existed in domains such as information retrieval for a very long time (i.e. TREC¹, Text Retrieval Conference). In image retrieval, benchmarks have started to develop from 2002, only. First benchmarks include TRECVID that started in TREC and has the goal to evaluate mainly video retrieval and ImageCLEF² that is part of the Cross Language Evaluation Forum (CLEF³). Since 2004, ImageCLEF has a medical retrieval task with the goal to evaluate medical visual information retrieval. An overview of the ImageCLEF medical task can be found in [12]. Once databases are available for participants, one of the hardest parts in organizing such a retrieval benchmark is the development of realistic search topics based on which the systems can be compared. For other domains the visual information retrieval behavior of users described in several articles, for example journalists searching for visual information to illustrate articles [9].

For ImageCLEF, several approaches of query topic development have been applied: In 2004, a domain expert defined search tasks that were typical for him; in 2005, topics were developed based on several surveys performed among image users [11], and in 2006, the topics were based on the log files of a medical image search engine on the Internet [10]. Topic development has several constraints. The goal is to have specific and non-ambiguous information needs and facilitate the job of relevance judges to judge for relevance or non-relevance. If the number of relevant images is too high, a comparison of several techniques becomes more difficult, because participating systems only submit a limited number of images and many images can lead to a strong variation by chance. On the other hand an information need has to correspond at least to a few relevant images in the database to allow for an evaluation at all.

These constraints make the topic development task often hard and time-consuming. Having an image retrieval system in clinical use and use such a system's log file would be the easiest option but currently no such system is available for us.

1. Methods

1.1. Base data used

The base data used includes a full day of queries performed at the Pubmed⁴ web site during 24 consecutive hours. The log file contained 2,689,166 queries. A more detailed analysis of this log file can be found in [8]. Most queries are in no connection with the visual field, and thus we discard them. Only information needs regarding visual content are of interest for us. Taking into account the large number of queries in the log file a manual analysis is not possible for the entire set but only for a subset of the data.

¹ <http://trec.nist.gov/>

² <http://www.imageclef.org/>

³ <http://www.clef-campaign.org/>

⁴ <http://www.pubmed.gov/>

1.2. Rules for filtering

Obtaining only image-related topics is difficult. The easiest way is filtering for queries in connection a modality. We filtered out queries containing the following key words: (image, video, media, xray, x-ray, CT, MRI, PET, tomography, ultrasound, ultrasound, endoscopy). Before applying these filters all plural forms were normalized to singular.

Among the frequent results of this filtering we made sure that at least two, or better three of the four following axes are fulfilled to result in topics that are specific enough:

- anatomic region;
- modality;
- pathology;
- visual observation (such as an enlarged heart).

Candidate topics that were obtained after these steps were subsequently used to perform test queries with several systems to make sure that at least a few results images are available in the data. Finally, the queries are ranked with respect to their visualness, meaning how much they seem to correspond to a visual search system or a text search system. Final goal is to develop 30 topics, ten of each category visual, textual, mixed.

1.3. Finding images corresponding to the query topics

In ImageCLEF, all query topics are created from a task description in three languages (English, French, German) with at least two accompanying images. After 30 query topics were identified, we needed to find at least 2 images per topic that correspond to the information needs and that can be used as input for visual search systems. We used Google image search to find such images and made sure that the copyright allows at least for a use in a non-commercial research environment. For a few of the topics it turned out to be hard to find query images and thus these queries were discarded.

2. Results

2.1. Most frequent queries

Table 1. The most frequent queries overall in the log file.

Query	Frequency
Finasteride	3601
Ibuprofen and toxicity not gastrointestinal	3421
One and a half syndrome	1751
#1 and #2	1242
Hypertension	801
Osteoclast tab12	767
Influenzae	765
Diabetes	640
Cancer	552
Heart	481

Table 1 shows the ten most frequent queries in the log file. It can easily be seen that none of the queries corresponds to a visual information need. This underlines the need for filtering. It also becomes clear that there is an extremely large number of different queries performed with Pubmed. Even the most frequent queries are rare compared to the overall number of queries performed. Most queries are very specific.

2.2. Image-related topics

Table 2. The most frequent queries containing topics related to visual information needs

Query	Frequency
MRI	58
Ultrasound	42
Otitis media	37
fMRI	33
Cardiac MRI	20
Endoscopy	20
Walsh CT	18
Lung ultrasound	15
Capsule endoscopy	15
Ultrasound for thyroid disorders	15

Table 2 shows the most frequently found query entries that contain the keywords that were chosen for filtering of visual topics. Again, it can be seen that only a few of them could actually be used for a retrieval benchmark making further processing necessary. This further processing is manual, going through the long list of visual queries and filtering out those containing at least two of the axes mentioned above. Frequencies of image-related terms are shown in Table 3.

Table 3. Overall number of queries with image related terms.

Term	Frequency
Image	1275
Video	298
Media	4774
Xray/x-ray	822
CT	230752
MRI	5578
Ultrasound	2140
Endoscopy	571

2.3. Final topics

Finally, 30 topics were distributed among the participants of ImageCLEF, divided into the categories visual, semantic, and mixed topics depending on whether they seem adapted for visual or textual retrieval systems. The classification was performed by a domain expert familiar with both techniques.



</TOPIC>
 <ID>S9</ID>
 <EN-DESCRIPTION>medial meniscus MRI</EN-DESCRIPTION>
 <DE-DESCRIPTION>MR des medialen Meniskusses</DE-DESCRIPTION>
 <FR-DESCRIPTION>IRM du ménisque interne</FR-DESCRIPTION>
 </TOPIC>

Figure 1. A visual query topic

Figure 1 shows an example for a visual topic and Figure 2 and example for a semantic topic more aimed at textual search engines. It can be seen that in the first case results images are expected to be rather homogeneous whereas in the second case the visual variety among the results can be extremely high.



```
<TOPIC>
<ID>s9</ID>
<EN-DESCRIPTION>pulmonary embolism all modalities</EN-DESCRIPTION>
<DE-DESCRIPTION>Lungenembolie alle Modalitäten</DE-DESCRIPTION>
<FR-DESCRIPTION>Embolie pulmonaire, toutes les formes</FR-DESCRIPTION>
</TOPIC>
```

Figure 2. A semantic query topic

2.4. Outcome of the benchmark

38 research groups from all continents and over 20 countries registered for the medical retrieval task and obtained access to the data and query topics. 13 groups finally submitted results that were compared. In total, 149 runs were submitted in the benchmark and evaluated by the organizers.

Several performance measures were calculated and compared to show various aspects of the systems. Lead performance measure is as in many other benchmarks Mean Average Precision (MAP). For the first time in 2007, best results were obtained by a fully textual retrieval system. On the other hand, for the first time, a fully visual system had extremely good results. This shows that mixed media retrieval has still a very large potential and that much can still be learned in this domain (combining the best visual and textual techniques did lead to a significantly higher result). Mixed media approaches also had by far the best results in early precision, which is most often the performance measure really important for a system user. More on the outcome of the medical benchmark in 2007 can be found in [13].

3. Conclusions

Benchmarks have become an extremely important part of many research domains, particularly in fields where basic research is transforming into applied research and where several prototypes are available. In these fields, a benchmark can focus research work into certain directions and objectively compare several techniques on the same bases, something that is otherwise most often impossible.

ImageCLEFmed has shown its importance through a large participation from worldwide research groups that is increasing every year. Only through realistic search

topics that correspond well to the available databases it is possible to motivate a large number of participants as they need to see an important advantage to participate in such an event and spend their time on the proposed tasks. The evaluation of a log file of the Pubmed search engine allowed us to develop 30 search topics that were subsequently used to compare the results of the participating groups. It is clear that most searches in such a log file are not for visual content, and even those information needs that contain visual requirements are often too vague to be used directly. Much manual intervention is necessary to obtain realistic and usable information needs. Tests with the data base are necessary to make sure that images corresponding to these needs are present.

In the end, such realistic query topics can help to create acceptance for visual retrieval techniques in medical institutions by showing their performance. Good techniques can be identified and compared with other techniques based on realistic tasks. Once first systems are in routine use in clinical institutions, their log files can even lead to topics better-adapted to real visual information needs.

References

- [1] AWM. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain, Content-Based Image Retrieval at the End of the Early Years, *IEEE Trans on Pattern Anal Mach Intell* **22** (200), 1349-1380.
- [2] H Müller, N Michoux, D Bandon, A Geissbuhler, A review of content-based image retrieval systems in medicine – clinical benefits and future directions, *Int J Med Inform* **73** (2004), 1-23.
- [3] TM Lehmann, MO Güld, C Thies, B Fischer, K Spitzer, D Keysers, H Ney, M Kohnen, H Schubert, BB Wein, Content-based image retrieval in medical applications, *Methods Inf Med* **43** (2004), 354-361.
- [4] P. Korn, N. Sidiropoulos, C. Faloutsos, E. Siegel, Z. Protopoulos, Fast and effective retrieval of medical tumor shapes, *IEEE Trans Knowl Data Eng*, **10** (1998), 889-904.
- [5] LHY Tang, R Hanka, HHS Ip, A review of intelligent content-based indexing and browsing of medical images, *Health Inform J* **5** (1998), 40-49.
- [6] C-R Shyu, CE Brodley, AC Kak, A Kosaka, AM Aisen, LS Broderick, ASSERT: A physician-in-the-loop content-based retrieval system for HRCT image databases, *Comput Vis Image Underst* **75** (1999), 111-132.
- [7] AM Aisen, LS Broderick, H Winer-Muram, CE Brodley, AC Kak, C Pavlopoulou, J Dy, CR Shyu, A Marchiori, Automated storage and retrieval of thin-section CT images to assist diagnosis: System description and preliminary assessment, *Radiology* **228** (2003), 265-270.
- [8] JR Herskovic, LY Tanada, W Hersh, EV Bernstam, A day in the life of Pubmed: analysis of a typical day's query log, *J Am Med Inform Assoc*, **14** (2007), 212-220.
- [9] M Markkula, E Sormunen, Searching for photos - journalists' practices in pictorial IR, *The challenge of image retrieval*, Newcastle upon Tyne, UK, 1998.
- [9] H Müller, C Boyer, A Gaudinat, W Hersh, A Geissbuhler, Analyzing web log files of the Health On the Net HONmedia Search Engine to define typical image search tasks for image retrieval evaluation, *Medinfo* **12** (2007), 1319-1323.
- [10] W Hersh, J Jensen, H Müller, P Gorman, P Ruch, *A qualitative task analysis for developing an image retrieval test collection*, International Workshop on Image and Video Retrieval Evaluation, Vienna, 2005, p. 11-16.
- [11] W Hersh, H Müller, J Jensen, J Yang, P Gorman, P Ruch, Advancing biomedical image retrieval: development and analysis of a test collection, *J Am Med Inform Assoc* **13** (2006), 488-496.
- [12] H Müller, T Deselaers, E Kim, J Kalpathy-Cramer, TM. Deserno, P Clough, W Hersh, Overview of the ImageCLEFmed 2007 medical retrieval and annotation tasks, *Lect Notes Comput Sci*, 2008 - in press.

Voice-controlled Data Entry in Dental Electronic Health Record

Miroslav NAGY^{1,a}, Petr HANZLICEK^a, Jana ZVAROVA^a, Tatjana DOSTALOVA^b,
Michaela SEYDLOVA^b, Radim HIPPMAN^b, Lubos SMIDL^c, Jan TRMAL^c and Josef
PSUTKA^c

^a *EuroMISE Center, Department of Medical Informatics, Institute of Computer Science
AS CR, v.v.i., Prague, Czech Republic*

^b *Department of Paediatric Stomatology, 2nd Medical School, Charles University in
Prague*

^c *Department of Cybernetics, University of West Bohemia in Pilsen*

Abstract. The EuroMISE Center focuses on new approaches in the field of electronic health record (EHR). Among others, the structured health documentation in dentistry in the form of an EHR is being systematically studied. This paper describes the evolution of the EHR developed at the EuroMISE Center named MUDRLite and its graphical component for dentists called DentCross. The summary of features of the DentCross component is followed by a brief description of automatic speech recognition and an ASR module. The problems with data insertion into EHR during examination of a dental patient lead to further research in the area of the automatic speech recognition in medical practice. Cooperation of engineers, informaticians and dental physicians resulted in an application called DentVoice which is a successful application of the ASR module and the DentCross component of the MUDRLite EHR. The junction of voice control and graphical representation of dental arch makes hand-busy activities in dental praxis easier, quicker and more comfortable. This will result in a better quality of the data stored in a structured form in dental EHR, thus enabling better decision making and use of decision support systems.

Keywords. automatic speech recognition, electronic health record, dental medicine

Introduction

Patients' health data are an integral part of dental healthcare documentation and they are constantly becoming more important. Besides insurance companies, information about patient's health status is required in practice in order to improve health care quality and decision making. There are still some obstacles that have to be surpassed.

One of them is the lack of rich structured data in dental medicine. Information still appears in the form of free-text-based documentation. Data structuring is on a relatively low level, mostly limited to tables, containing more or less standardized symbols (e.g.

¹ Corresponding Author: Miroslav Nagy, EuroMISE Center, ICS AS CR v.v.i., Pod Vodarenskou vezi 2, 182 07 Prague 8, Czech Republic; E-mail: nagy@euromise.cz

"/" for caries, "-" for pulpitis, or "x" for a tooth to be extracted), placed in the section corresponding to a particular tooth. The main disadvantage of using such a system is the loss of detailed information on the localization, the size and the character of hard dental tissues defects. Further information on an oral cavity is not sufficiently structured as well. Information concerning changes of oral cavity mucosa, periodontitis, orthodontic anomalies, preventive oncological examinations, etc. is described in a limited space of one line of the form or together as other findings in the form of a free text.

It has been a common situation in dental practice that during check-up the dentist dictated patient's data to a nurse, who filled out paper based forms. Nowadays, computerization in dental practice brought major changes in data entry. Nurse enters patient's data into a specialized computer application. As an example of a dental application used in Czech Republic we can mention the PC Dent software by DialogMIS [1]. The PC Dent is a medical information system for dentists, containing textual patients' records, support of scheduling and a graphical tool for dental status check-up and treatments. The PC Dent is one of the most elaborated medical software systems in dentistry available on the Czech market and is received quite well by its users.

Entering data into a computer during the examination of a patient requires either costly changing of sterile gloves, which is also time-consuming, or the participation of a nurse, entering the dictated data. These problems could be possibly overcome by use of the automatic speech recognition in EHR. Voice commands usage has been examined since 1990's [2], [3] as a convenient replacement of computer keyboard and mouse. Necessity of using human voice to control a computer or other device arose in typical hands-busy environments such as surgery [4] or dentistry [5].

1 DentCross Component and Automatic Speech Recognition in MUDRLite EHR for Dentistry

1.1 MUDR EHR and Dental Knowledge Base

Several European projects, international standards and long-lasting cooperation with physicians helped us to develop a pilot EHR system called MUDR (MULTimedia Distributed Record) [6], [7]. It is based on the three-tier architecture with an unusual data-storing approach based on two main structures described by tools of a graph theory, named knowledge base and data-files.

In the years 2004 and 2005 the dental knowledge base was created. Dentists and informaticians from the joint workplace of the EuroMISE Center produced a model in the form of a knowledge tree as a part of the knowledge base of the MUDR EHR. The knowledge tree comprised of basic information about patient, family history, social history, personal history, information about medication and exhaustive information about patient's oral cavity status from the dentist's point of view. Approximately 1000 dental concepts were structured during this process. The knowledge tree served as a model basis for further dental EHR created at the EuroMISE centre.

1.2 MUDRLite

The MUDRLite is an EHR inspired by the MUDR EHR. Its architecture is based on two layers. The first one is a relational database. Currently, MS SQL server versions 7 and 2000 are supported. The second layer is a MUDRLite User Interface (MUDRLite UI) running on a Windows-based operating system using Microsoft .NET Framework.

The database schema corresponds to the particular needs and varies therefore in different environments, as opposed to the fixed database schema in the MUDR data layer. MUDRLite universality is based on a different approach. The core of MUDRLite – MUDRLite Interpreter – is able to handle various database schemas. This feature often simplifies the way of importing old data stored with other databases or files.

All the visual aspects and the behaviour of the MUDRLite UI are completely described by an XML configuration file. The end-user can see a set of forms with various controls placed on them by appropriate XML elements. MUDRLite operates as a kind of commands' interpreter; it processes the instructions encoded in the so called MLL Language as described in [8] and manipulates the database layer as well as the visual aspects of the MUDRLite UI.

1.3 DentCross Component

To gain MUDRLite's user-acceptance in the field of dental medicine we have developed an advanced component representing the dental cross, which is a crucial part of healthcare documentation in dental medicine. Its development was motivated by clinical practice in the dental medicine domain. The DentCross component was included dynamically into the user interface in a form of user-defined custom component.

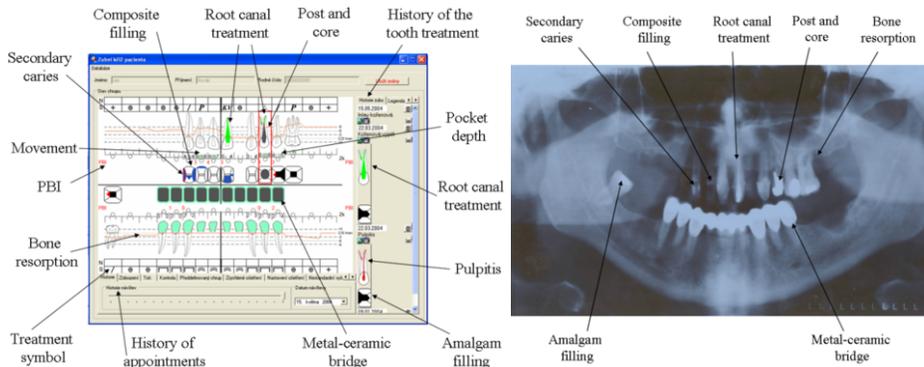


Figure 1. The DentCross Component – a Case Study Example and the Corresponding Tomograph Picture.

The DentCross component is implemented as a stand-alone library DentCross.dll, completely developed for the .NET Framework platform using the Microsoft Visual Studio .NET 2003 development tool.

A user-defined component is inserted by the custom element of the MLL Language with the following mandatory attributes: "dll" specifying the name of the

assembly the component is implemented in, and “class” specifying the name of the main class of the included component.

For the end user, the DentCross component looks like a kind of a dental panoramic tomograph. This component is fully interactive and enables to record fully structured dental medicine information that can be inserted user-friendly by mouse or keyboard. A dentist can choose among about 60 different actions [9], treatment procedures or tooth parameters that are displayed graphically and lucidly.

The Figure 1 depicts a screenshot of the DentCross component showing dental status of a case study patient and X-ray image representing the same situation. Apparently, the graphical representation of the dental arch produced by DentCross component is as illustrative as the X-ray image and acceptable by the patient. Modern dental chairs have an integrated monitor that can be viewed by both the dentist and the patient, thus improving mutual trust and as a result lessening patient’s stress.

1.4 Automatic Speech Recognition Module

The automatic speech recognition (ASR) engine was implemented by the Department of Cybernetics, University of West Bohemia in Pilsen [10]. The module was implemented as a standalone application, running in server-mode in background.

The server communication protocol is proprietary, running on the top of TCP/IP stack. The communication protocol enables the startup and shutdown of the recognition process, the run-time configuration of the recognition task and also receiving the recognized phrases by a client.

The ASR system is speaker-independent and is based on a statistical approach. The goal was to implement a recognition module featuring high accuracy yet able of very fast operation. It comprises a front-end, an acoustic model, and a decoding block as can be seen on Figure 2.

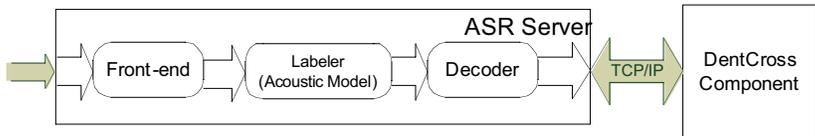


Figure 2. Automatic Speech Recognition functionality.

1.4.1 Front-end

The speech signal is digitized at 8 kHz sample rate and converted to the u-law format. Then the pre-emphasized acoustic waveform is segmented into 25 milliseconds frames with 15 ms overlap. A Hamming window is applied to each frame and static PLP cepstral coefficients (PLP_CCs) are computed. Moreover, the delta (first order derivatives) and the delta-delta (second order derivatives) coefficients are computed and appended to the feature vector.

1.4.2 Acoustic Model

The acoustic model is based on modeling of triphones. Each individual triphone is represented by a three state left-right HMM (Hidden Markov Model) with a continuous output probability density function assigned to each state. Each density function is approximated as a mixture of multivariate Gaussians with diagonal covariance matrices.

The number of mixture components for each state was obtained experimentally. Since a variety of noise sounds (e.g. loud breath, click on the microphone and noise of an audio channel) can appear in every utterance, a set of noise HMM units was developed and trained.

1.4.3 Decoder Block

The decoder uses a Viterbi search technique together with efficient beam pruning [11]. The search is performed on a crossword context dependent HMM state network. The state network is generated from EBNF (extended Backus–Naur form) metasyntax notation described in JSGF (Java Speech Grammar Format) format. The whole recognition network is constructed from one or more together connected grammars.

To improve the robustness, the decoder contains a rejection technique to refuse incoming utterances that are not accepted by the grammar (or grammars). In this case the ASR engine informs the client that the utterance was not successfully recognized.

2 DentVoice Application

Interconnection of the ASR module and the DentCross component of the MUDRLite EHR resulted in an application called DentVoice. Junction of voice control and graphical representation of dental arch makes hand-busy activities in dental praxis easier, quicker and more comfortable. Dentists were involved in the whole development process; therefore manipulation with the DentCross component was designed to be as easy as possible.

The prototype application consists of a DentCross component with integrated TCP/IP client of the ASR server and a voice commands definition file. The ASR client uses a DentCrossHandler class that implements all functionality of the DentCross component.

The speech recognition is activated immediately after DentCross component start-up. The recognition process can be paused or stopped by a special voice command or using the user interface. Voice commands can be divided in two groups: global manipulating commands and context dependent commands.

Global commands are always available and are designed to manipulate the recognition process (pause, resume and stop) and to close message boxes opened by the application to warn the user.

Context dependent commands rely on the current state of the DentCross component and can be further divided into 33 command groups because it can fall into one of 33 states (e.g. tooth treatment, caries placement, caries type, root canal treatment material).

3 Discussion

First experience with the DentVoice application showed that the need of abbreviated form of commands is inevitable. Since the main goal of the voice controlled approach is to make the data entry easier and faster, further shortcuts should be found. On the other side, some minimal length of commands should be kept, as the length of commands (the longer the better) plays significant role in improving the recognition hit

rate. Abbreviated forms of positions on the tooth such as D for distal or B for buccal, are very similar from the phonetic point of view and therefore easily mistakable by the ASR module. Further improvement of the accuracy of the ASR module has been achieved by having voice commands separated in state dependent groups, thus limiting the number of possible commands in a particular state.

Forthcoming development regards usage of computer synthesized speech. The DentCross will read the actual status of the patient's teeth stored in the database and the dentist will just check if the information corresponds with reality. If so, the check-up ends without any necessary action, otherwise, the reading process is paused and the dentist dictates the divergence he or she found. In order to continue the check-up the reading process is resumed. The synthesized computer voice will serve as a feedback to the dentist that can stay focused and look in patient's mouth only to validate the correctness of the stored data.

DentCross component has been successfully used in forensic dentistry [12] to improve identification of disaster victims. Voice controlled version of the DentCross, DentVoice application, will be getting further improvement in data entry during the dead corpses inspection.

Acknowledgement

This paper was supported by the project 1ET200300413 of the Academy of Sciences of the Czech Republic.

References

- [1] Dialog MIS, <http://www.dialogmis.cz/english/index.html> (last accessed November 6, 2007).
- [2] Grasso MA. Automated speech recognition in medical applications. *MD Computing* 1995; 12(1):16-23.
- [3] Grasso MA, Ebert D, Finin T. Acceptance of a speech interface for biomedical data collection. *Proc AMIA Annu Fall Symp* 1997; 739-43.
- [4] Reichenspurner H, Damiano RJ, Mack M, Boehm DH, Gulbins H, Detter Ch, Meiser B, Ellgass R, Reichart B. Use of the voice-controlled and computer-assisted surgical system ZEUS for endoscopic coronary artery bypass grafting. *J Thorac Cardiovasc Surg*, 1999; 118: 11-16.
- [5] Drevenstedt GL, McDonald JC, Drevenstedt LW. The role of voice-activated technology in today's dental practice. *J Am Dent Assoc*, 2005; 136(2): 157-161.
- [6] Hanzlicek P, Spidlen J, Nagy M. Universal electronic health record MUDR. *Stud Health Technol Inform* 2004; 105: 190-201.
- [7] Hanzlicek P, Spidlen J, Heroutova H, Nagy M. User interface of MUDR electronic health record. *Int J Med Inform*, 2005; 74(2-4): 221-7.
- [8] Spidlen J, Hanzlicek P, Zvarova J. MUDRLite: Health record tailored to your particular needs. *Stud Health Technol Inform* 2004; 105: 202-9.
- [9] Spidlen J, Pies M, Teuberova Z, Nagy M, Hanzlicek P, Zvarova J, Dostalova T. MUDRLite – an electronic health record applied to dentistry by the usage of a dental-cross component. *IFMBE Proceedings Vol.11*; 2005.
- [10] Muller L, Psutka J, Smidl L. Design of speech recognition engine. *Lect Notes Comput Sci* 2000; 1902: 259-64.
- [11] Psutka J, Muller L, Matousek J, Radova V. *Mluvíme s počítačem česky*. Praha, Academia, 2006.
- [12] Zvarova J, Dostalova T, Hanzlicek P, Teuberova Z, Nagy M, Pies M, Seydlova M, Eliasova H, Simkova H. Electronic health record for forensic dentistry. *Methods Inf Med* 2008; 47: 8-13.

A Large, High Resolution Tiled Display for Medical Use: Experiences from Prototyping of a Radiology Scenario

Bernt Ivar OLSEN^a, Sanjaya Babu DHAKAL^a, Odd Petter ELDEVIK^b,
Per HASVOLD^c, Gunnar HARTVIGSEN^{ac}

^a*Department of Computer Science, Faculty of Science, University of Tromsø*

^b*University Hospital of Northern Norway*

^c*Norwegian Centre for Telemedicine*

Abstract. The scope of the project was to assess the value of using high-resolution, very large displays in a hospital setting. We applied a scenario informed prototyping method and user-involvement in order to do this. Initial results suggest that the technology could prove very useful in clinical conferencing settings like the communicating process between the radiology department and the other hospital departments using their services. The possibility of bringing more visual information simultaneously to the audience is especially intriguing. However, issues such as floor control – who administers the (extra) information space and information-overload, are imminent in interface design and our prototype suggests that the clinicians do want functionality that stresses these issues.

Keywords: human computer interaction, large, high-resolution displays, radiology, user informed design

1. Introduction

The renaissance of very large, high resolution and (relatively) low-cost displays have been foreseen and expected for some years already. The transition from CRT monitors to LCD display panels and plasma screens with increasingly more pixels and larger size have to some degree already met those expectations, and video projectors are also reaching ever new heights in terms of image quality and resolution. However, even though technology, like large high-resolution plasma screens and LCDs have been introduced into clinical environments, like surgery, radiology and others (i.e. [1]), technology like a display wall (see Figure 2) has not been studied in such an environment before. The affordances of large, high-resolution displays have been studied for some years already [2], but not in terms of clinical use.

Jonathan Grudin stated in his 2001 study of multi-monitor use [3] that a typical large display only occupies about ten percent of your visual field we have when only moving our eyes. In the natural world we probably feel quite restricted by such a

narrow field of view. Large, high resolution displays allow us to both enlarge the image viewed and to be able to move close to it – without losing detail as you would with a regular projector. Recent research suggests that large display surfaces do provide certain benefits for the users [4],[2]. However, the large physical display provides an excellent common workspace for people to collaborate as well. A fair body of research within this context has been done in the last decades (see. i.e. [5] for an overview on this topic) and it is within this paradigm we believe that the major benefits from large displays truly lies. We have used a user-centered approach, including observations of users in real work activities, to inform the creation of a crude interface prototype for selected clinical conference situations.

2. Methods and materials

In our study we have used a scenario-based prototyping approach combined with observations of work activities at the Radiology Department to inform design and possible use of large, high-resolution displays in teamwork settings at a hospital.

The display wall is created from 28 projectors and an equal number of cluster nodes running the RedHat Linux operating system. The 28 projectors create, through hardware and software technology, a more or less seamless image that is 7168x3072 pixels – approximately 22 megapixels (see Figure 2). The Display Wall is approximately 220' across.

We conducted some initial observations at the University Hospital of Northern-Norway (UNN). Based on the findings from those observations, one case was selected as the basis for design of a prototype. User-involvement was invoked on the level that the users were introduced to the capabilities of the display wall and the purpose of the demo was discussed. The Radiology department, being the selected case, then chose a patient case for the prototype and the image material along with textual information noted on the patient case was handed to the design team. Imaging material that was selected by the Radiology Department the patient case, consisting of DICOM formatted images and textual information was used to create a prototype interface for the display wall. We conducted a demo of the prototype where potential benefits of using a high resolution display like the display wall in a medical setting at the Radiology Department was highlighted and discussed. The main objective was to assess the qualities of a large high resolution display wall as a collaborating mechanism in the context of a typical radiology demonstration setting (scenario). Hence, the demo was thoughtfully planned to provoke discussion and reflection on what such technology has to offer. We videotaped the demo and comments were transcribed. Results from this demo are our primary findings.

In designing the prototype we had image material like CT, MR, CT-angio stitched together to a large enough image to fit the wall (22 mega pixels). For this purpose we used ImageMagick¹ and borrowed the familiar interface from the current image viewer at the Radiology Department as a “frame” around the images (see Figure 2). We used a parallel image viewer, developed at the Computer Science department, capable of viewing images across the cluster of computers running the display wall in order to have a very responsive interface that do not require several seconds of loading when zooming and panning along the large images on the wall.

¹ www.imagemagick.org

To describe the prototype we built we use the framework that Houde and Hill presents in [6]; role, look-and-feel and implementation. Our prototype is developed along all the three dimensions. The role of the artifact is investigated by regarding the display wall as the communicating medium for the radiology team and staff from any other department in the hospital. Look and feel component is medium to high fidelity and resolution. We provide an actual looking interface on a real display wall. On the implementation side, this is a prototype with limitations in fidelity and resolution of the interface elements. We developed the scenario subsequent to the observations done at UNN. The scenario was not explicitly shared with the user group (radiologists), but was for internal use among the designers. The scenario: “A Radiologist is presenting a patient case to another department. The latter department has requested the images. The radiologist is using a 22 megapixel; the equivalent of 28 ordinary displays.” It is a very ambiguous and general scenario, although it describes the specific use of the artifact.

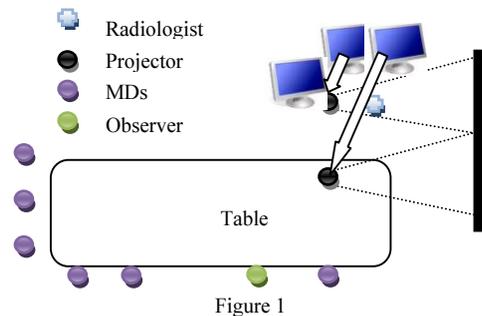
3. Findings

3.1. Findings from the observation

The observational studies performed were mainly done in order to look at the presently available equipment in the hospital and only to a limited degree to focus on how the tasks are actually performed. The first observation was performed at an oncology department at UNN. We observed a morning meeting at the oncology department and a weekly teleconferencing-educational seminar between the current Tromsø department and an off-site part of this department located in the city of Bodø². Using a display-wall for this scenario, we could i.e. expect a more sense of presence from the remote site at the local site having the display-wall

[7] in the video-conference. However – we decided cooperatively with the oncology department not to pursue this scenario as a candidate for our prototype. The second observation was at a conference room located at the Pathology department of the University Hospital. We were given a demo of the equipment and an unstructured interview with one of the leading pathologists at the department was

conducted. The conference room we were shown had room for approximately 40 people and is typically in use for conferences on for instance brain tumors, urology, mammography cases and gynecology. The equipment used in this facility was a 2-projector lineup with two screens, where one screen typically displays variable content depending on the case studied while the right displays the pathology image being selected by the person handling the microscope. We agreed that the pathology conference room was interesting for our scenario and that the pathologist would provide input to the prototype and be present at the demonstration. The third observation session was at a radiology meeting with a group of neurologists. Based on



² The oncology department is one administrative unit, but is physically divided by 1000km

the previous meetings and the discussions we had both before the meetings, during, and afterwards we already knew that the clinicians are used to large displays and a two-projector setup. That meant that we could advance somewhat and start looking for particular needs for the kinds of features the display-wall technology offers. We were looking for situations where additional information about patient was requested and where image material was browsed through, for instance where the method of scrolling through image material was used. Figure 1 shows a sketch of the setup at the radiology board. Notice the very typical setup of three computer screens for a radiology station and the two ceiling-mounted projectors used to show the images from the patient and communicate findings. The arrows from the computer screens illustrates that the image from these screens are shared to the audience. Typical here is that the imaging data is the only one being shared visually. If any other data is requested it is typically read out loud by the radiologist controlling the workstations PCs. A typical patient presentation proceeds like this: (1) the radiologist reads from the patient journal/charts relevant information. (2) Images are loaded and selected to be put onto the projector screens, (3) the patient and the findings are discussed in the group. The discussion is in general involving 2-4 people including the radiologist.

3.2. Findings from the Demo/Focus group

The patient selected by the radiologists was one with substantial image material of several modalities. The patient data was anonymized and transferred in its original DICOM format. The demo of the prototype was an intimate setup at the display wall lab at the computer science department. Present were the design team, the current head of the radiology department, a former head of the radiology, now head of clinical IT-systems at UNN, a visiting senior radiologist and a chief physician at the pathology department. The demo itself was a series of interface examples taking advantage of the large display area. In their comments on the prototype they are generally comparing it to the two-projector setup in Figure 1.

The participants liked the extra space and immediately suggested that patient information and history could be available. They suggested that the display should be divided into three parts, where one of the three could be dedicated patient information, while the two remaining could be used to imaging, as they are today. Figure 2 shows pictures from the demo of our prototype. 2 a) shows a complex of images and image-modalities including pathology 2 b) shows 4 selected images from the image set on the left in the image, where one physician moves closer to study details. One of the main hypothesis before the demonstration of this prototype was that the ability to view more information simultaneously would be the main interest in the large display. This is the reason for stitching many CT and MR images together on the different slides. However, even though facilitating the overview- and detail concept proved to be highly successful among the audience, the displaying of many images at the same time seems to provoke the conception that this is a step backwards – that this is what everyone thought when digital imaging was introduced with PACS some years ago. One of the comments during the demo illustrates this point:

“ ... historically, in the development of PACS, one thought that one would recreate what we had with analogue x-rays, where we had large displays with many images hanging in a rack. The bigger surface on which to hang images, the better it was – we thought. In the first implementations of PACS they thought that the more images they could view – on as many

displays as possible, the better. (...) Then the breakthrough came, when someone was thinking simple and noticing that ‘you never look at more than one image at a time. You only need the option of choosing one image out of several and look at that more closely’ ”.

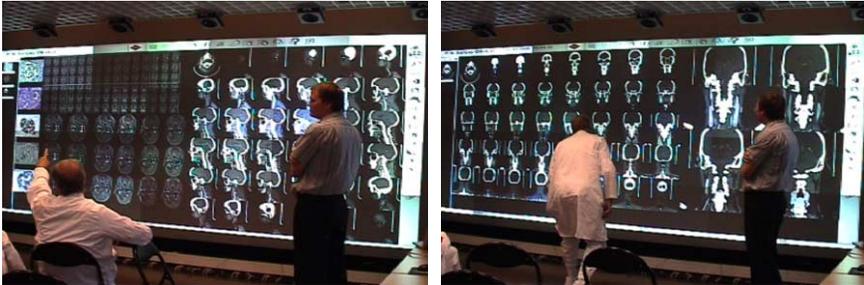


Figure 2

Furthermore, one of the participants felt he was “drowning in information” when all the images were sequenced over the entire wall. Another comment was that it could be useful to have certain kinds of information available all the time on the screen (static), while the rest could change. An example of static information was patient information like name, address and which department the patient belonged to.

As for the collaboration aspect of the large display there was a couple of issues that came up during the presentation. One was tied to the fact that the potential “added” information pieces that would be on the wall is not currently shared – or at least not visually with the audiences in such settings. As mentioned before, the textual information, patient history, etc. are currently shared only orally and so the presenter has the possibility of “screening” this information before it reaches the audience (selecting and omitting pieces information as he/she finds relevant). Furthermore, this material is not authored today with visual sharing in mind, and changing this might also change how the content is created. This might for instance mean that whom ever is presenting might use longer time in preparing (and in the first place – writing) this information for presentation. This is, of course, related to how the presentations are prepared today – which needs to be studied. The other related issue was that of floor control: who selects and manipulates the displayed information on the wall? A quote from one of the participants:

“ (...) if you take mammography, for instance – there are three things in use: x-ray images, EPR and pathology. Is it so that one person can control the x-ray images, one can control the EPR and a third can control the pathology? It has to be like that (...)”

The issue of whom and how one controls the presentation and what goes onto the screen is an important one, and a real challenge for the interface design. There were also quite a lot of questions regarding the tiling of the screen and whether this could be better done in a finished system. The audience seemed to agree that the difference in coloring between the projectors would have to be reduced.

4. Conclusions, discussion and further work

Although information overload seems to be the basic concern that is provoked by our prototype, we believe that this issue needs to be counter-balanced by the fact that the large display technology is not - in this context - meant to be used by a single individual alone. The large surface and high resolution provides a novel situation where much more information could be made visible simultaneously. We have experienced that the re-introduction of cascading images on the large display may not be particularly useful, even if it is a tempting way of utilizing the extra display area. Instead, the cues are thus far, that the extra screen real estate should be configured so as to bring more diverse information that facilitates the collaborate problem-solving and – decision making process that clinical conferencing is about – not more of the same information. A consequence of this matter is the problem how to highlight what information is in focus at any time – or how to select or “frame” the relevant information at any time [8]. The first and most obvious thought is to present *overview* (i.e. a collection of the generally perceived most important, diverse pieces of information on the patient) by default and to enable selection of *detail* on an ad-hoc basis. The effects of the large, high-resolution displays remains to be seen in real use by clinicians, which we hope to achieve through iterative development of a radiology interface for the wall and subsequent opportunities for clinical testing.

5. Acknowledgements

We would like to thank the technical staff and the SHARE group at the Department of Computer Science and professor Otto Anshus in particular.

6. References

- [1] A.N. Averbukh, *Assessment of a Novel, High-Resolution, Color, AMLCD for Diagnostic Medical Image Display: Luminance Performance and DICOM Calibration*. Journal of digital imaging, **16**(3): 2003. p. 270.
- [2] M. Czerwinski. *Toward Characterizing the Productivity Benefits of Very Large Displays*. in *INTERACT 2003*. Zurich, Switzerland: IOS Press. 2003.
- [3] J. Grudin, *Partitioning digital worlds: focal and peripheral awareness in multiple monitor use*, in *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2001, ACM Press: Seattle, Washington, United States.
- [4] D.S. Tan, et al., *Physically large displays improve performance on spatial tasks*. ACM Trans. Comput.-Hum. Interact., **13**(1): 2006. p. 71-99.
- [5] K. O'Hara, M. Pery, and E. Churchill, *Public and Situated Displays: Social and Interactional Aspects of Shared Display Technologies (Cooperative Work, 2)*: Kluwer Academic Publishers. 2004.
- [6] S. Houde and C. Hill, *What do Prototypes Prototype?*, in *Handbook of Human-Computer Interaction*, S. Helander, et al., Editors. Elsevier Science: Amsterdam. 1997.
- [7] D.S. Tan, et al., *With similar visual angles, larger displays improve spatial performance*, in *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2003, ACM Press: Ft. Lauderdale, Florida, USA.
- [8] N. Lund, et al., “*Watch the Document on the Wall!*” *An Analytical Model for Health Care Documents on Large Displays*, in *Web Information Systems Engineering – WISE 2007 Workshops*. 2007, p. 395-406.

The Contextual Nature of Usability and its Relevance to Medical Informatics

Dag SVANÆS^{ab}, Anita DAS^b, Ole Andreas ALSOS^a

^a*Department of Computer and Information Science, NTNU, Norway*

^b*National EHR research center, NTNU, Norway*

Abstract. We report from three usability evaluations of health information systems that illustrate the value of seeing usability as a context dependent property of a product. We show how the definition of usability in ISO 9241-11 can be used as a guiding principle in the evaluation, specification and design of such systems. The contextual view on usability is particularly important for health information systems because of their great diversity concerning user groups, tasks, and work environments.

Keywords. Human factors, user-computer interface, standards, health information systems, usability, context-of-use

Introduction

Compared to other fields of informatics, medical informatics has been relatively late in adopting a focus on usability. This can be seen both as a drawback and as an opportunity. A positive consequence is that the field can draw on 25 years of research in human-computer interaction (HCI). One important sign of a mature research field is that it has developed a shared set of well-defined concepts. Within HCI, one of the most central concepts is *usability*. Since its infancy in the early 80s, the HCI field's understanding of usability has changed from a focus on efficiency and "user-friendliness" to a mature definition made explicit in the standard ISO 9241-11 [1]. An important shift in the understanding of usability has been from seeing it as a property of the software as such to see it as a relation between the software and its *context of use*. This has important consequences for how software should be designed, evaluated and implemented. The heterogeneity of the use situation for health information systems makes awareness of the contextual nature of usability particularly important for such systems. We will illustrate the relevance of a contextual perspective on usability for health information systems with findings from three recent research projects.

1. Related Work

Already 15 years ago Jakob Nielsen [2] discussed the relation between usability and the overall *system acceptability* of a product. He argued that other factors such as price,

compatibility, and social acceptability are important. More recently, it has been argued that aesthetics and “desirability” are just as important as usability [3].

It has been empirically shown that the *technology acceptance* of a product can be predicted with technology acceptance models such as UTAUT [4]. Such models list usability (“ease of use”) as one of the most important factors affecting the user’s choice, thus showing the importance of a focus on system usability.

The definition of usability in ISO 9241-11 has been the basis for ISO 25062 [5], which specifies a common industry format for reporting the results from usability tests.

2. Usability Defined

ISO 9241-11 defines the usability of a product as “*the effectiveness, efficiency, and satisfaction with which specified users achieve specified goals in particular environments*”. The ISO standard defines usability in terms of three measures: *effectiveness, efficiency* and *satisfaction*, and three aspects of the use situation: the *users, their goals* and the *use environment*. The use situation as a whole is referred to as the product’s *context of use*.

Effectiveness is the most important measure. It is the extent to which the goals of the users are achieved. In usability tests this is often measured as task completion. *Efficiency* is about the resources necessary to achieve the goal. In usability tests this is often measured as completion time. *Satisfaction* is about the user’s subjective assessment of the product. In usability tests, user satisfaction is often measured with standardized questionnaires. Together, these three measures give an indication of how “usable” the product is. Experience from empirical research shows that these measures can not be reduced to one single measure, and that all three should be reported when usability tests are done [6].

The ISO definition of usability differs from many other ISO definitions by being context dependent. It says that the usability of a product cannot be measured as a property of the product as such, but only in relation to a specific *context of use*. This is different from context independent definitions as the Kilogram. The weight of a laptop PC is independent of its user, its purpose of use and its use environment.

The ISO definition of usability refers to three aspects of the use situation: (1) *Specified users* require that the users are identified. For usability evaluations of systems in actual use, the specified users are the actual users. For usability tests it is necessary to characterize the future users to such an extent that it is possible to select a representative set of test subjects. (2) *Specified goals* are about the users’ purpose when using the product. In usability tests, the users’ goals are defined through the tasks that the test subjects are asked to carry out. (3) *The environment* is the physical and social conditions of the use situation. This includes physical aspects, such as furniture and noise levels, and ergonomic aspects such as placement of equipment. In usability tests it is necessary to imitate the environment to the necessary level of detail.

3. Usability is Contextual

Let us illustrate the ISO standard’s definition of usability with an example from medical informatics. Many electronic health record (EHR) systems currently offer access through wireless personal digital assistants (PDAs). Let us focus on such a

system's function for retrieving lab results for a patient. For the current purpose, let this function be what the ISO standard refers to as the *product*. The *effectiveness* of this product is to what extent the users (i.e. the doctors and the nurses) are able to get access to the lab results for a given patient. Given that the users are able to complete their task, the *efficiency* is about how much effort they have to put in to achieve it. The efficiency can be quantified as time used on the task, need to consult written material, and help from colleagues. One way of measuring the *user satisfaction* could be to ask the users questions about how they like the system on a scale from 1 to 5.

Let us assume that the *specified users* for our hypothetical EHR system are doctors and nurses. This would then exclude patients, hospital administrators, visitors or any other person present in a hospital who might want to look at lab results. If a patient picks up the PDA, fails to get access to his lab results, and complains about the usability of the product, he has misunderstood what usability is all about. This product was not intended for patients, and it is therefore meaningless to talk about its usability for users outside of its intended user group. Our hypothetical product could have been designed to allow for access by patients, but that would have made it a different product with a different intended user group, and consequently with different criteria for evaluation of usability.

A similar argument goes for the *specified goals*. If a user criticizes the system for not allowing for text messaging, that is not a statement about the usability of the EHR lab result function, but a wish for some new functionality. Such statements are of course legitimate, and feedback of this sort should even be encouraged in the dialogue with the users, but they say nothing about the usability of the product.

Let us, in a similar fashion, assume that the intended *use environment* for our hypothetical product is specified to be within the walls of the hospitals and in good light conditions. The fact that the PDA will be useless in the badly lit back of an ambulance far away from the hospital wireless network is consequently irrelevant in a discussion about the usability of this product, as long as ambulances are not included as one of its intended use environments.

4. Context-of-use Illustrated with Three Cases

We will illustrate the relevance of a contextual perspective on usability from three recent usability studies performed in our usability laboratory. The usability laboratory at the Norwegian National EHR Research Center (NSEP) in Trondheim has an 80 m² room with movable walls that allows for full-scale simulations of situations from hospitals. That makes usability tests possible with a number of users simultaneously.

4.1. Case A: Getting the User Group Right

In cooperation with the system vendor, we did a usability test of an eHealth service for primary care that allows patients to communicate with their doctor through a web interface. The web service allows the patients to make appointments, send requests to their doctor, and ask for renewal of prescriptions.

One of the most critical usability problems found was related to SMS codes sent by the system to verify the identity of the user. These codes had to be read from the patient's mobile phone and typed in on the PC. The codes were random sequences of digits and letters, e.g. "Bq7cP8". For a number of users this became a major obstacle.

We observed that this was particularly true for users with pain and on strong medications. This observation led us to question the validity of the usability test, as one might expect that a considerable number of eHealth users will be pain patients. To test if pain and medication actually has an effect on the patient's ability to use the system, we did a controlled experiment where a group of cancer patients (N=14) and a matched control group (N=14) were exposed to the eHealth system. The study showed a statistically significant difference in completion rate between the two groups [7].

The lesson learned from this case is that one has to be very precise in the definition of the user groups for health information systems. Large between-subject variations in test results are indications that one might have to split the user group into subgroups.

4.2. Case B: Getting the Physical Environment Right

A number of new hospitals now install bedside terminals for the patients. Such terminals are currently to a large extent used for entertainment and web browsing. In cooperation with one of the vendors of these terminals, we explored the potential for letting doctors use handheld devices (PDAs) as input device for the bedside terminals. Seven different prototype user interfaces were implemented. They were tested on a scenario where a doctor uses a bedside terminal to show X-ray images to a patient [8].

Due to patient safety and privacy issues we were not allowed to test the prototypes *in situ*. The usability tests were done in our lab with a replication of a patient room with a hospital bed, a touch screen and a PDA. The tests were run with pairs of doctors and patients. After having tried out all versions, the doctors and patients were asked to rank the different solutions by sorting cards representing the alternatives. They were asked to give reasons for their ranking. Figure 1 shows to the left the recorded video from the test, and to the right a combined video with the bedside terminal and the PDA.



Figure 1. The doctor uses his PDA to select which X-ray image to show to the patient.

An analysis of the factors that influenced their ranking showed that the usability of the different design solutions were to a large extent not a result of the usability of the Graphical User Interfaces (GUIs) as such, but of how each design solution affected the dialogue between the physician and the patient. These factors included free eye contact, the ability to have a shared view, and distance to the screen. A “desktop” usability test, focusing only on the system's GUI, would for this case have given misleading results.

The lesson learned from this case, is that for usability tests of health information systems it is necessary to imitate the environment to a high degree of realism.

4.3. Case C: Getting the Scenario Right

We are currently involved in the usability testing of a new version of a hospital EHR system with new modules for computerized physician order entry (CPOE) and lab result access. The current version only works on stationary computers, while the new version will also work on laptop PCs with WLAN access. The version to be tested is in a beta release, and due to patient safety and privacy issues we are not allowed to test it *in situ*. Five persons were involved in each test: a doctor, a nurse and three actors simulating patients.

To ensure the realism of the usability test, a portion of a ward has been replicated in our lab, including meeting room, corridor, patient rooms and patient beds. To get the scenario right, it was necessary to do field studies of actual work practice and include health workers in the specification of tasks and goals. It was also necessary to give the patients in the tests realistic medical histories, so that the doctors and nurses had goals that mimicked those in real life. The first phase of the project was a baseline test where pairs of doctors and nurses performed a pre-round, round and medication scenario with the current combination of stationary EHR system and paper-based ordering and lab results. Figure 2 shows the wall setup of the laboratory and a scene from the test.

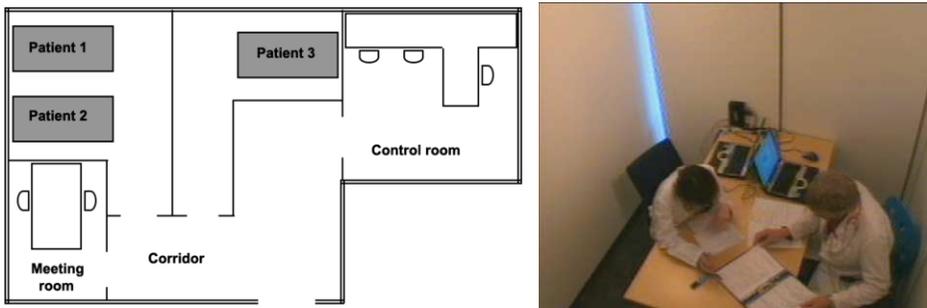


Figure 2. Laboratory configuration and a scene from the usability test.

To check the validity of the baseline test, the video recordings from the test were compared with field notes from an actual ward. The comparison revealed a number of differences related to the scale of the social unit being tested. We had not included any interruptions in the scenario, while the field notes showed that medical work to a large extent is event driven, with many interruptions and with many tasks going on in parallel. We used this knowledge in the planning of the next phase of the test, by introducing unexpected interruptions in the scenarios. Interruptions included physicians being paged due to emergency and nurses being called to other patient rooms.

The lesson learned from this case is that medical work is very complex, and to get valid test results the use scenarios must replicate the tasks and goals of the health workers in great detail. Domain experts should be included in the design of the scenarios, and tests should be compared with observations from real life.

5. Discussion and Conclusions

Our experience from doing usability tests of health information systems shows the importance of taking into consideration the contextual nature of usability as it is defined in ISO 9241-11. The medical domain is highly specialized, with a number of different professions and patient groups, a large number of different activities, and many different work environments – all within the walls of the same hospital. Some systems have only one user group, while many health information systems have a number of user groups, often with different tasks and work environments. It is consequently meaningless to talk about the usability of a particular information system without being very specific about its intended user groups, tasks, and work environment.

This has consequences for the usability evaluation of such systems, but also for the specification and design phase. Due to the complexity of the *context of use* for such systems it is important that user groups, tasks and work environments are defined in great detail before a new system is being designed. If one or more of these factors are not well defined, the development project runs the danger of making a product that might have a high usability for the project team's imagined context of use, but not for the actual context of use.

The Human-Computer Interaction field has produced a number of design techniques, such as contextual inquiry [9], personas [10], and scenario-based design [11] that can help in the identification of the correct context of use. Properly used, such techniques can enable developers of health information systems to not only “make the thing right”, but also to “make the right thing”. For this purpose we suggest that the ISO definition of usability can be a powerful guiding principle.

References

- [1] ISO/IEC, ISO 9241-11 *Ergonomic requirements for office work with visual display terminals (VDTs). Part 11 Guidance on usability*, ISO/IEC 9241-11:1998 (E). ISO 1998.
- [2] Nielsen, J., *Usability Engineering*. 1993: Academic Press.
- [3] Norman, D.A., *Emotional Design: Why We Love (Or Hate) Everyday Things*. 2004: Basic Books.
- [4] Venkatesh, V., et al., *User Acceptance of Information Technology: Toward a Unified View*. MIS Quarterly **27** (2003), 425-478.
- [5] ISO/IEC, ISO/IEC 25062. *Software Engineering - Software product Quality Requirements and Evaluation (SQuaRE)-Common Industry Format (CIF) for Usability Test Reports*. ISO/IEC 25062:2006 (E). ISO 2006.
- [6] Hornbæk, K. and Law, E.L.C., *Meta-analysis of correlations among usability measures*. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2007: p. 617-626.
- [7] Das, A., Svanæs, D., and Faxvaag, A., *Cancer patients' ability to use an online health service: a usability study*. In preparation.
- [8] Alsos, O.A. and Svanæs, D., *Interaction techniques for using handhelds and PCs together in a clinical setting*, in *Proceedings of NordiCHI 2006*. 2006, ACM Press: Oslo, Norway.
- [9] Wixon, D.R., Ramey, J., and Holtzblatt, K., *Usability in practice: field methods evolution and revolution*, in *CHI '02 extended abstracts on Human factors in computing systems*. 2002, ACM: Minneapolis, Minnesota, USA.
- [10] Pruitt, J. and Grudin, J., *Personas: practice and theory*, in *Proceedings of the 2003 conference on Designing for user experiences*. 2003, ACM: San Francisco, California.
- [11] Carroll, J.M., *Making use: scenarios and scenario-based design*, in *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*. 2000, ACM: New York City, New York, United States.

OPTISAS a new method to analyse patients with Sleep Apnea Syndrome

Adrien UGON^{ad}, Carole PHILIPPE^a, Slawomir PIETRASZ^a,

Jean-Gabriel GANASCIA^d, Pierre P LEVY^{abc1}

^aHôpital Tenon(Assistance Publique Hôpitaux de Paris), France ;

^bINSERM, UMR-S707, Paris, F-75012;

^cUniversité Pierre et marie Curie –Paris 6, UMR S 707, Paris, F75012

^dLIP6, Université Pierre et marie Curie –Paris 6, Paris, F75015

Abstract: OPTISAS is a visualization method that allows describing very precisely a patient with Sleep Apnea Syndrome. Using the events scored by the physician, our method gives a set of graphs that are a detailed representation of the condition, sleep stage and position, in which the events occur. This helps for the diagnosis. This is possible thanks to the application of Generalized Caseview method. The method proceeds in two steps, defining the reference frame and using this reference frame to visualize data. The reference frame is built by using a supin/unsupine binary criterion, a six type event criterion and a sleep stage ordinal criterion. The main result is the visualization of the indexes (average number of events by hour) associated with the events. This allows a more accurate diagnosis showing the precise influence of the position and of the sleep stage on the events.

Keywords: Generalized Caseview; Information visualization; Sleep Apnea Syndrome; Polysomnography; Diagnosis; Signal Processing

1. Introduction

Sleep Apnea Syndrome (SAS) is a real public health problem. The sleep apnea – hypopnea syndrome occurs in 2-4% of the middle aged population [1], causing daytime sleepiness, cognitive deficits and road traffic accidents. Its association with increased cardiovascular [2] and cerebrovascular morbidity [3] has been clearly recognized. Young and colleagues estimated that 2% of women and 4% of men meet the criteria for the clinical syndrome of sleep-disordered breathing. Caring of these patients at an early stage is very important for treatment efficiency and complication prevention. The main existing exploratory means is the polysomnography. This examination is an overnight multichannel recording dedicated to the accurate diagnosis of sleep disorders. Due to

¹ Address for correspondence

Pierre Lévy

Hôpital Tenon, Public Health Department

4, rue de la chine, 75970 Paris cedex 20, France

Email : pierre.levy@tnn.aphp.fr

their multimodality, the generated data are complex to analyze, forming a typical relevant problem for information visualization methods application.

The method used is the Generalized Caseview method (GCm). This method is a pixelization method [4] that has proven its efficiency in processing complex data [5-7]. The point of this paper is to show the contribution of the GCm to polysomnographic characterization of the SAS. As a matter of fact this method allows the OPTImisation of the care of patient with SAS (OPTISAS). In the methods we will explain the main parameters of a polysomnography and the basis of the GCm, then in the result the main images obtained when applying the method to the data of a patient will be presented. Finally an evaluation of the method will be proposed thanks to its application to a set of patients.

2. Methods

The polysomnography (PSG) is the medical examination used to detect the sleep disorders. It consists of simultaneously recording many neurological and respiratory parameters, during a night. Among the sleep disorders, we can define some breathing events [8]. The apnea is a clear cessation of breathing during sleep lasting at least ten seconds. The hypopnea is a clear amplitude reduction of a validated measure of breathing during sleep (between thirty and fifty percents from baseline) that is associated with an arousal or an oxygen desaturation larger than three percents. If there is still an effort to breathe, objectified by an increase of the activity of the respiratory muscles, the apnea is obstructive; in the other case, it is central. If the apnea begins as a central apnea but towards the end there is an effort to breathe without airflow, the apnea is mixed. The results of the analysis of a PSG are given by indexes that are defined as the average number of events that occur every hour during the sleep. Thus, the Apnea-Hypopnea Index (AHI) indicates the severity of the Sleep Apnea Syndrome (SAS). But to make a diagnosis, the physicians need to see a temporal representation of the patient's night, with the sleep stages, the position (figure 2) and the scored events. Then, it is necessary to know the main respiratory events and their type. Each specific event is measured thanks to a specific index. Moreover, the duration of the events is also important, in particular the duration of the longest event. Two other events are important: those are the desaturations and the micro-arousals, which are measured thanks to their respective indexes.

There are still a large quantity of key numbers that are used by the physicians to make their conclusions and the diagnosis. They are not detailed here but it is important to know that they are presented in tables of numbers lying on ten A4 pages: the source data come from the polysomnographic report generated by Somnologica developed by Medcare [9].

The Generalized CaseView is a pixelization method [4,5,6,10] that gives a visual representation of data tables that point out groups very clearly; they are next very easy to identify. In that aim, the data are structured and localized on a 2D-map using three criteria: firstly, the data are separated in 2 groups using a binary criterion; each group is taking place in one side of the map. Then, they are separated in many groups, according to a nominal criterion; each group has to take place in a row on the graph. Each data is finally placed using an ordinary criterion. This gives the order of the data from the central axis allowing defining a symmetry. Every cell contains a data and its background is colored by defining a color scale (these cells are called "infxel" [4]).

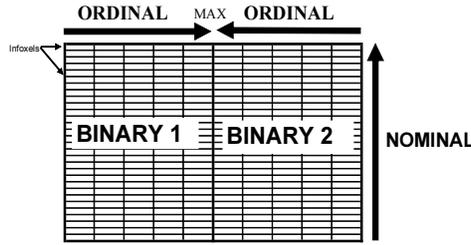


Figure 1 The three criteria of the reference frame of the GCm

Our method is to draw a set of graphs in order to visualize quickly all the information and have a precise conclusion and precise diagnosis, considering all the indexes given by position of the patient and sleep stages. Every graph shows the data, aggregated by the position and/or the sleep stage. The first graph shows all the data recorded during the night (Somnologica usual representation); the next graphs indicate the indexes of the events for each couple of position and sleep stage. In that purpose, we propose to use the generalized CaseView method. As binary criterion, we use the position, supine position versus unsupine position. The ordinary criterion is the sleep stage ordered by depth (S3-S4, S1-S2, REM). As nominal criterion we use all the scored events, which correspond to a respiratory or neurologic state.

3. Results

The approach of the method is to translate basic data into aggregated data. This translation allows extracting the semantic content of the data.

The first graph is a temporal representation of the events during the night. This graph is necessary to see the architecture and fragmentation of the sleep and to have a first idea about the type of events and their severity. In some case, it is possible to detect some correspondences between an event and a position and/or a sleep stage.

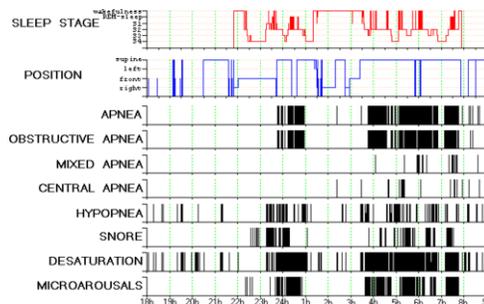


Figure 2 Usual representation of a polysomnographic data

Thus, the figure 2 shows that the patient has many obstructive apneas during sleep, a few hypopneas, associated to desaturations and microarousals. The patient has two periods of intensive apneas while he is sleeping in supine position. He changed of position many times during the night and awoke for 2 hours in the night. However he slept about 7 hours that is enough for giving conclusions.

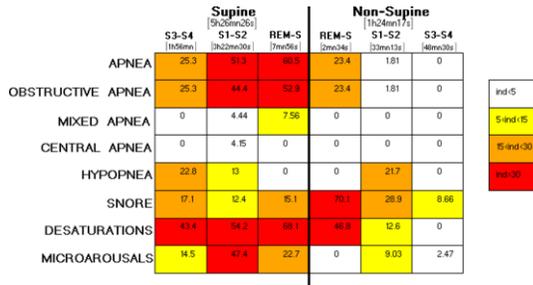


Figure 3 Visualization of the indexes using the generalized Caseview method (GCM)

Figure 3 was drawn using the *GCM*. The infoxels contain the indexes for each scored event in the corresponding position and sleep stage. The colour scale is shown on the right. We can see the indexes depending on the position and the sleep stage. It is clear that the patient has an obstructive apneas syndrome depending on the position. The dark color indicates that the SAS is severe. The apneas are associated with desaturations and microarousals. There are many apneas in deep sleep, which is not common.

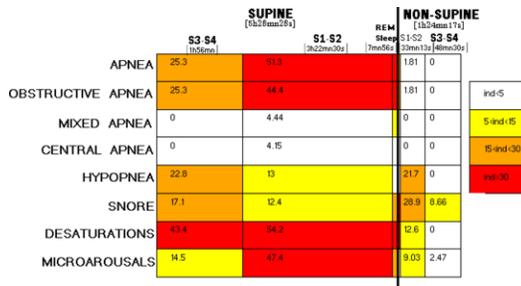


Figure 4 Visualization of the indexes using the GCM

Figure 4 is almost the same as figure 3, except the width of the columns is proportional to the time spent in these conditions. The patient clearly slept just a very few minutes in REM-sleep. This figure confirms the supine feature of the SAS.

Figures 5 and 6 allow completing the analysis by showing other parameters.

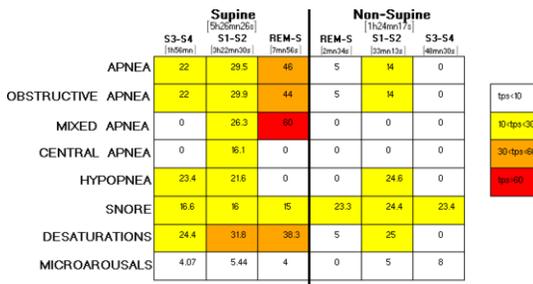


Figure 5 Visualization of the average durations of the events using the GCM

In term of average length, the figure 5 shows that the apneas are mostly not very long. We already know that apneas were more frequent in supine position. Figure 5 shows that they are also longer. It appears that in unsupine position, the patient has a normal profile, the disorders occur when the patient sleeps in supine position.

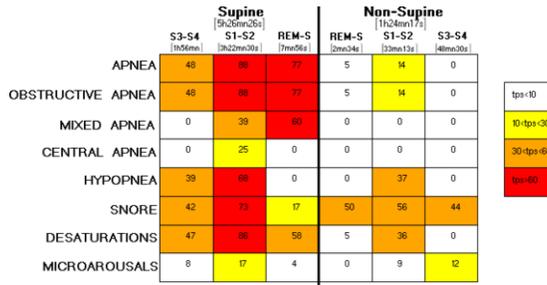


Figure 6. Visualization of the durations of the longest events using the GCm

The figure 6 represents the duration of the longest event for each condition. They are quite long. Some apneas last more than 60 seconds. Once again, in unsupine position, the events are not severe.

As review, the patient has a severe obstructive apnea syndrome. The main respiratory events are obstructive apneas. These apneas are related to supine position. The apneas are averagely not very long but some last more than 60 seconds, which is very long. Our method points even out that in unsupine position, the patient has a normal profile considering the frequency and the duration of the events.

4. Evaluation

The results were analysed by the chief physician of the sleep laboratory. We randomly chose a set of 13 patients with SAS. The expert had to give his conclusions concerning the existence of a link between the SAS and a sleep stage and/or a position by looking at our set of graphs. Our method was better for 7 patients considering the sleep stage and for 4 patients considering the position. Globally for 8 patients (62%), our method offered extra information.

5. Discussion

We studied other methods that analyse sleep data. The mainly used visualization of the PSG is a polysomnogram that is a temporal linear representation of all the recordings used in diagnosing sleep disorders.

Fernandez-Leal and Moret-Bonillo [11] proposed a new method based on temporal knowledge analysis of the PSG. Their method is an AI approach. The main interest seems to be that their method enables “the temporal quantitative, qualitative and causal constraints to be represented and processed”. An important result is like ours, a gain in diagnosis accuracy. Compared to our method, their method is analytical when ours is global, i.e. visual.

Fred et al [12] presented an expert system used to build the most plausible diagnosis in all sleep disorders, regarding polysomnographic and clinical data. In this tool, a graphical user-interface helps to control every step of the diagnosis. Like OPTISAS, the goal of the graphical representation is to resume the patient's data for the expert to understand what underlies the conclusions. The aim is even different as ours is to point out any dependences of the SAS to sleep depth or position, letting appear

some hidden aspects of the SAS that the patient suffers whereas Fred et al just illustrate all the rules that made the expert take his decision.

Guimarães et al [13] classified patterns by using several 2D and 3D graphical representations in the framework of Self Organizing Maps, processing then them with AI methods to discover hidden information in polysomnographic data. On the contrary, OPTISAS draws a simple image that is then analyzed by a human expert able to identify patterns describing accurately the patient.

OPTISAS mixes a simple processing with an accurate description of the patient.

6. Conclusion

Our method gives a brief but exhaustive synthesis of common indicators used to diagnose a SAS, letting know in particular very easily the type and the severity of the SAS but also any influence by a position or a sleep stage, what was confirmed in 62% of cases in the evaluation. Such a formalization of the data permits to have an accurate diagnosis and to adapt the treatment.

We are working on a more accurate and formal evaluation and validation of OPTISAS and we are implementing it in the daily activity of the physicians of the sleep laboratory. This study can thus be considered as the first step of a larger project that aims to establish groups of patients depending on the results obtained in our method.

- [1] Young T, Peppard PE, Gottlieb DJ, Epidemiology of obstructive sleep apnea: a population health perspective. *Am J Respir Crit Care Med*, 2002;165(9) : 1217-39.
- [2] Marin JM, Carrizo SJ, Vicente E, Agusti AG, Long term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure : an observational study. *Lancet* 2005 Mar 19-25; 365(9464) :1046-53
- [3] Arzt M, Young T, Finn L, Skatrud JB, Bradley TD, Association of sleep-disordered breathing and the occurrence of stroke. *Am J Respir Crit Care Med*, 2005 Dec 1 ;172(11):1447-51
- [4] Lévy PP Pixelization Paradigm : outline of a formal approach. In: Pixelization paradigm. Proceedings of the first Visual Information Expert Workshop (VIEW2006). Lecture Notes in Computer Sciences. Editors PP Lévy et al. Vol 4370. pp.3-11. Springer. 2007.
- [5] Lévy PP The case view a generic method of visualization of the case mix. *International journal of Medical Informatics* 2004, 73, 713-718.
- [6] Lévy PP, Duché L, Darago L, Dorléans Y, Toubiana L, Vibert J-F, Flahault A. ICPCview : visualizing the international Classification of Primary Care. *Connecting Medical Informatics and Bio-Informatics. Proceedings of MIE2005*. R. Engelbrecht et al. (Eds). IOS Press, 2005. pp623-628. 2005
- [7] Xhemo E, Fajac A, Boire J, Levy PP. MDRView: a visualization of the polymorphisms of MDR1 (ABC1) gene in breast cancer. *Conf Proc IEEE Eng Med Biol Soc*. 2007;1:4592-4594.
- [8] Kushida CA, Littner MR, Morgenthaler T, Alessi CA, Bailey D, Coleman J Jr, Friedman L, Hirshkowitz M, Kapen S, Kramer M, Lee-Chiong T, Loube DL, Owens J, Pancer JP, Wise M. Practice parameters for the indications for polysomnography and related procedures: an update for 2005. *Sleep* 2005 Apr 1;28(4):499-521.
- [9] <http://www.embla.com/support/downloads/download.asp?hFile=1060&cbLanguages=4>
- [10] Lévy PP. Le case view une methode de visualisation du case mix. *Journal d'économie médicale* mars 2002; 20(2), 118-127.
- [11] Fernandez-Leal A, Moret-Bonillo V, Temporal knowledge in the intelligent diagnosis of sleep Apnea syndrome: a New Approach.. proceedings of the 29th annual international conference of the IEEE EMBS, August 23-26 2007, Lyon, France.
- [12] Fred A, Filipe J, Partinen M, Paiva T, In T.Paiva and T.Penzel, editors, *European Neurological Network*, number 78 in Studies in Health Technology and Informatics, chapter 3.3, pages 127 -147. IOS Press, 2000.
- [13] Guimarães G, Peter J-H, Penzel T, Ultsch A: A method for automated temporal knowledge acquisition applied to sleep-related breathing disorders; *Artificial Intelligence in Medicine* 23(3) :211-237 (2001)

7. Learning, Modelling and Simulation

This page intentionally left blank

Predictors of Preterm Birth in Birth Certificate Data

Karen L. COURTNEY^{a1}, Sara STEWART^b, Mihail POPESCU^c, and Linda K. GOODWIN^d

^a *University of Pittsburgh, USA*, ^b *Independent Researcher, USA*, ^c *University of Missouri – Columbia, USA*, ^d *Duke University, USA*

Abstract. Demographic factors have been shown to be moderate predictors of preterm birth in prior studies which used hospital databases and epidemiologic sample surveys. This retrospective study used de-identified 2003 North Carolina birth certificate data ($n=73,040$) and replicated the statistical and computational methods used in a prior study of an academic medical center's data warehouse. Receiver Operating Characteristics (ROC) curves were used to compare results across methods. Due to differences between the data collected for birth certificates and the original clinical database, five of the seven demographic variables in the clinical database model were available for model testing (maternal age, marital status, race/ethnicity, education and county). Even with a reduced model, multiple methods of statistical and computational modeling supported the earlier findings of demographic predictors for preterm birth. The reduced model AUC results were acceptable (logistic regression = 0.605, neural networks = 0.57, SVM = 0.57, Bayesian classifiers = 0.59, and CART = 0.56), but lower than in the prior study as might be expected for a reduced model. On a population level, these results support a prior demographic predictor preterm birth model generated from a clinical database and the use of computational methods for model formation. Additional testing for stronger predictor models within birth certificate data is suggested as birth certificate data is a parsimonious population dataset already routinely collected.

Key Words. Premature Birth, Health Services Research, Modeling, Obstetrics & Gynecology, Nursing

Introduction

Worldwide, preterm birth and/or low birth weight contribute to 24% of neonatal deaths [1]. In the United States, preterm births (birth prior to 37 weeks gestation) are the leading cause of neonatal deaths not related to congenital birth defects [2]. Rates of preterm birth have been increasing in the United States and in 2003, the national average was 12.7% of all births within the United States [3], but rates among some racial and ethnic populations such as African Americans (18.3%) in comparison to among non-Hispanic white mothers (11.4%) [3].

¹ Corresponding Author: Karen Courtney, School of Nursing, University of Pittsburgh, 415 Victoria Building, Pittsburgh, PA 15261, USA; Email: Karen.courtney@alumni.duke.edu

Preterm infants who survive are at increased risk for developmental disabilities, cerebral palsy, mental retardation, blindness, chronic lung problems, and other lifelong conditions which add to the pain and suffering families may experience with the birth of a preterm baby. The annual financial toll for initial hospital care of preterm infants is estimated to be US\$15.5 billion [2] each year, with additional costs for children with special needs going well beyond this already astronomical amount. Small increases in gestational age are also associated with lower mean hospital charges. In one study, the mean hospital charges for full term infants was US\$4,788 compared to US\$10,561 (33 – 36 weeks), US\$55,792 (29-32 weeks), and US\$239,749 (26 – 28 weeks) for preterm infants [4]. Clearly, strategies are needed to find solutions to this important preterm birth problem. Complicating the search for effective preterm birth prevention interventions is a lack of an adequate and reliable preterm birth prediction model.

Exact etiology of spontaneous preterm birth is not known at this time and is likely to be the interaction of multiple risk factors [2, 3]. The literature is replete with contradictory research findings that yield minimal guidance for perinatal practitioners, and there is no scientifically validated framework for preterm risk factors. As an example, despite the prior widespread use of a preterm risk scoring tool, this tool was later found to be ineffective in accurately identifying most preterm births [5].

More recent modeling for preterm birth prediction has used medical records to investigate the role of personal health behaviors, history and demographics in preterm birth. However this approach has also resulted in inconsistent results. Goodwin, et al. [6] used multiple data mining techniques to study 19,970 racially diverse pregnant women and found seven demographic variables produced 0.72 area under the curve (AUC). It is interesting to note that in contrast to their results using seven variables, using all 1,622 variables available in the dataset produced only a slight increase in AUC (0.04) when predicting birth outcomes. In using receiver operating characteristic (ROC) techniques, chance prediction would be 0.50 AUC and perfect prediction would be 1.0, so 0.72 is a respectable prediction that suggests demographic data are important but, obviously, improved predictive accuracy is needed for clinical decision support.

Other studies have supported several of the demographic variables noted in the Goodwin model including: maternal age; marital status race and/or ethnicity [7]. Other research has also found personal income and neighborhood income characteristics to be important demographic predictors of preterm birth [8].

Sampling in these prior studies potentially could have affected the resulting preterm prediction models. Researchers have noted considerable variation in socio-demographic variables as well as preterm birth outcomes in comparison of a study sample drawn from local clinics and the local population characteristics [9]. As the Goodwin et al. [6] model was based on the medical records of academic medical center patients during years 1986 to 1995, this sample may have also affected the resulting prediction model. Patients receiving care in an academic medical center may represent higher acuity patients and not be representative of the population as a whole. Further testing of this prior model with either a broader sample or with population records was needed.

1. Methods

1.1. Research Design

This study was a retrospective, secondary analysis of North Carolina birth record data. The purpose of this study was to examine if a demographic preterm prediction model previously generated by data from academic medical center's clinical database could be applied to a population database. Following approval by the Health Sciences Institutional Review Board, we attempted to replicate a prior demographic preterm birth prediction model [6] with birth certificate data. Variables within the birth certificate data were as closely matched with the variable definitions from the prior model as possible. Two variables, mother's religion and payor source were not available within the birth certificate file and therefore omitted from the replication model. Preterm birth for this study was a dichotomous variable defined as a birth prior to 37 weeks gestation.

1.2. Sample and Data Source

The data source for this project was the publicly-available, de-identified, birth certificate records for North Carolina for 2003 [10]. These data were the most recent data available at the beginning of the study. The North Carolina birth records were chosen as the prior study was based on medical records from an academic medical center located in North Carolina, but this study's dataset (2003) did not contain any records from the previous study's birth dataset (1986-1995) [6]. This study's data set contained 120,168 live births and included maternal, paternal, infant and health care system variables. As state law requires birth certificate completion for all births, this data set is estimated to represent 99% of all births [3]. The exact number of undocumented births is unknown though likely small in number. When filtered for out-of-state births, induced or stimulated labor or multiple births, the data set contained 73,040 birth records.

1.3. Analysis

This study replicated the statistical and computational methods used in the prior study: logistic regression, neural networks, classification and regression trees (CART) and also included Support Vector Machines (SVM) and Bayesian classifiers for model comparison. This project used the Matlab® Neuronal Networks package and the SVM classifier implemented in the Matlab® Statistical Package. The Bayes classifier was coded in Matlab® by one of the authors (MP). Receiver Operating Characteristics (ROC) curves were used to compare results across methods.

Logistic regression was chosen based on the primarily categorical data being used. The response variable was binary (preterm or not preterm). With the exception of age, all the data were categorical. From the birth certificate data, the variables were categorized in the same manner as the original study. For example, in the birth certificate data, education is listed as the number of years of education. We put these education data into categories to match the variables used by the previous study. Our raw data also had ethnicity and race as distinct categories, which was not done in Goodwin et al.'s study [6]. Therefore, we had to combine these two variables from the birth certificate dataset in order to match original study definitions.

The NN structure consisted in N input binary inputs, 2N hidden layers and 2 output nodes, one for each class (preterm or term). The number of inputs, N, varied according to the number of variables used and the discretization of the continuous variable such as maternal age. For the case when N was large (for example, the county variable) we used principal component analysis (PCA) to reduce the dimensionality of the input space. The training of the NN was performed on 2,000 records from the 2004 North Carolina birth certificate data.

The Bayes classifier was implemented in a simple fashion by using the class mean. An unknown pattern was assigned to the class with the closest (using Euclidean distance) mean. More sophisticated Bayesian approaches did not seem to improve the classification result.

ROC curves are used to give a graphical representation of results of prediction. It is a graph of sensitivity versus specificity (true positives vs. true negatives) given a binary classifier. The area under the curve (AUC) represents predictive accuracy of the model. At 0.5, it means the results can be explained by chance. The ROC curve was useful for comparison in this study, as we had a binary classifier (preterm vs. not preterm) and all the methods gave us an AUC value. Even with the AUC values, it is helpful to compare the visual representation of the curves, which is done in Figure 1.

2. Results

Table 1: Comparison of Results across Studies

Method	Original Study (Goodwin, et al., 2001) AUC Results (n = 19,970)	Current Study AUC Results (n = 73,040)
Logistic regression	0.66	0.605
Neural networks (NN)	0.64	0.57
CART-based Custom Classifier Software (Goodwin, et al., 2001)/Classification & Regression Trees (CART)	0.72	0.56
Rule induction	0.67	Not used
Bayesian classifiers	Not used	0.59
Support Vector Machines (SVM)	Not used	0.57

Due to differences between the data collected for birth certificates and the original clinical database, five of the seven demographic variables in the clinical database model were available for model testing (maternal age, marital status, race/ethnicity, education and county). Even with a reduced model, multiple methods of statistical and computational modeling supported the earlier findings of demographic predictors for preterm birth. The reduced model AUC results were acceptable (logistic regression = 0.605, neural networks = 0.57, SVM = 0.57, Bayesian classifier = 0.59, and CART = 0.56), but lower than in the original study.

Figure 1 compares the ROC curve results among the different statistical and computational analyses. As illustrated, the results of this model are consistent across multiple statistical and computational models.

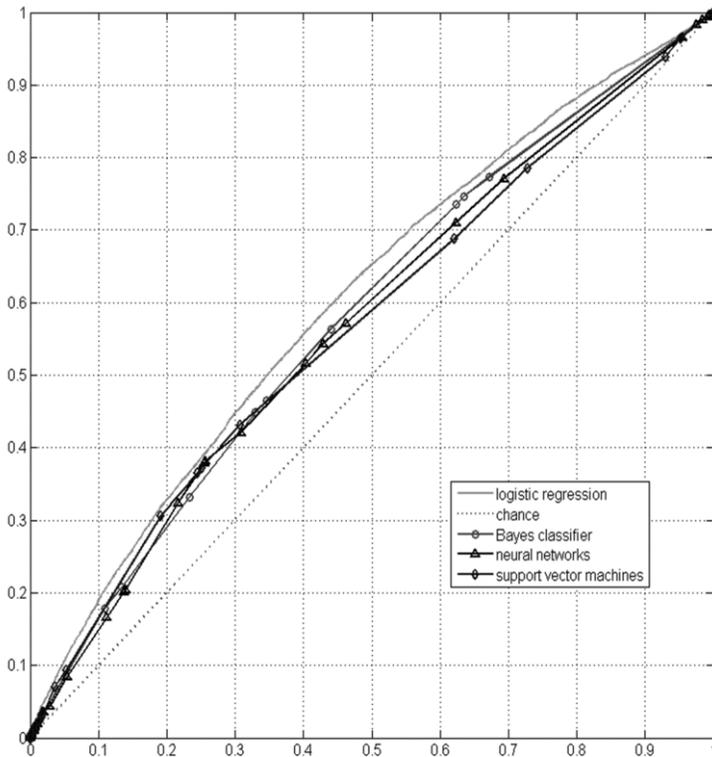


Figure 1: Comparison of Receiving Operator Characteristic (ROC) Curves by Computational and Statistical Methods

3. Discussion

The results of this study yielded lower AUC results than the original work; however this was not unexpected as two of the original variables in the model were unavailable in the new dataset. Although the direction of these results do support the original demographic model proposed by Goodwin et al. [6], these results also suggest that further refinement of this socio-demographic model is needed. The birth certificate data set contains other variables not available to the original research team in the clinical database. These variables include paternal demographic factors, maternal health behaviors and maternal medical history which may be important and should be incorporated into future socio-demographic model research. Of particular note, the

birth certificate data set contains measures of prenatal care adequacy: the Kotelchuck Index and the Kessner Index which were not available in the prior clinical dataset. Previous research has suggested that adequacy of prenatal care can affect preterm birth outcomes [11]. These measures should be tested in further refinement of a socio-demographic preterm prediction model.

4. Conclusion

The significance of demographic predictors in a pregnant population center around opportunities for intervention where socio-economic and geography factors can be mediated to prevent preterm birth and other pregnancy complications. Refinement of a socio-demographic prediction model using existing population data sources is the next step. Future models can be used in the design of clinical decision support systems. Additional research will need to examine how well subsequent socio-demographic models predict birth outcomes geographically in order to identify communities most at risk and target interventions effectively.

Acknowledgements:

This work was supported in part by the National Library of Medicine Biomedical and Health Informatics Research Training Grant T15-LM07089-14.

References:

- [1] World Health Organization. Chapter 2: Making Pregnancy Safer. HRP Biennial Report 2000-2001 2001 1/11/2007 [cited 1/11/2007]; Available from: http://www.who.int/reproductive-health/publications/biennial_reports/2000-01/Chapter_2.PDF
- [2] March of Dimes. Perinatal Overview. 2005 [cited 06/07/2005]; Available from: www.marchofdimes.com/peristats/
- [3] Martin JA, Hamilton BE, Sutton PD, Ventura SJ, Menacker F, Munson ML. Births: Final data for 2003 Nat. Vital Stat. Rep.. Hyattsville, MD.: Nat. Center for Health Statistics; 2005. Report No.: Vol 54 (2).
- [4] Cuevas KD, Silver DR, Brooton D, Youngblut JM, Bobo CM. The cost of prematurity: Hospital charges at birth and frequency of rehospitalizations and acute care visits over the first year of life. *AJN*. 2005;105(7):56-64.
- [5] Creasy RK. Preterm birth prediction: where are we? *American Journal of Obstetrics & Gynecology*. 1993;168:1223-30.
- [6] Goodwin LK, Iannachione MA, Hammond WE, Crockett P, Maher S, Schlitz K. Data Mining Methods Find Demographic Predictors of Preterm Birth. *Nursing Research*. 2001;50(6):340-5.
- [7] Rolett A, Kiely JL. Maternal sociodemographic characteristics as risk factors for preterm birth in twins versus singletons. *Paediatric and Perinatal Epidemiology*. 2000;14(3):211-8.
- [8] Luo Z-C, Kierans WJ, Wilkins R, Liston RM, Mohamed J, Kramer MS, et al. Disparities in Birth Outcomes by Neighborhood Income: Temporal Trends in Rural and Urban Areas, British Columbia. *Epidemiology* November. 2004;15(6):679-86.
- [9] Savitz DA, Dole N, Kaczor D, Herring AH, Siega-Riz AM, Kaufman J, et al. Probability samples of area births versus clinic populations for reproductive epidemiology studies. *Paediatric and Perinatal Epidemiology*. 2005;19(4):315-22.
- [10] North Carolina State Center for Health Statistics. IRSS Data Services. 2006 [cited 05/01/2006]; Available from: <http://www.irss.unc.edu/ncvital/index.html>
- [11] Taylor CR, Alexander GR, Hepworth JT. Clustering of U.S. Women Receiving No Prenatal Care: Differences in Pregnancy Outcomes and Implications for Targeting Interventions. *Maternal and Child Health Journal*. 2005 2005/06//;9(2):125-33.

The p53 Network Modeling – Current State and Future Prospects

Raul Florin HORHAT^{a,1}, Gheorghe I. MIHALAS^a and Mihaela NEAMTU^b

^a*University of Medicine and Pharmacy Timisoara, Romania*

^b*West University of Timisoara, Romania*

Abstract. p53 gene is a central hub in a network that have the role to protect the cells against carcinogenesis. A lot of research work has been done in this field, a huge amount of data was collected and many experiments have been performed. For a better understanding of the processes and for the help of experiments design some mathematical models were proposed in the literature. In this paper we make a brief overview of these models, pointing out their strong points and weak points. Some improvements of the models were also presented with some graphical representations of numerical simulation. Finally some conclusions were made regarding future prospects of p53 network modeling.

Keywords. Mathematical models in medicine, Human genome related, Modeling, Numerical simulation, p53.

Introduction – biological aspects

P53 gene was identified in 1979 by Arnold Levine, David Lane and Lloyd Old, but it took ten years after, until 1989, to be found its key role in the cell as tumour suppressor gene. From that moment, the research activity grow up very quickly, in other ten years were published more than 17,000 articles and over 10,000 tumour-associated mutations in p53 have been discovered, in organism ranging from humans to clams. In this moment, a Google search returns more than 8,600,000 results related to p53.

The gene's name, P53, is due to its protein molecular mass: it is in the 53 kilo Dalton fraction of cell proteins. The gene also gets a surname as "Guardian of the genome" due to its vital role that it plays in protecting the cell against tumor degeneration.

P53 activity can be described as the main defense mechanism against genetically mutations. The gene starts its job only when DNA damage occurred, otherwise it is in dormant stage [1]. Its main two actions are related with its protein concentration in the cell nucleus and here there are: a low or a brief increase in concentration determines cell cycle arrest, but a high or prolonged elevation of P53 concentration shifts the cell to apoptosis [2]. Taking into account these two opposite outcomes it is now clear that

¹ Corresponding Author: University of Medicine and Pharmacy Timisoara, Splaiul Tudor Vladimirescu, nr. 14, 300173, Timisoara, Romania, E-mail: rhohat@umft.ro

the P53 concentration should be kept in tight control. Here appears mdm2 gene that is in charge for this, resulting a negative feedback loop between p53 and mdm2 [3].

It was noticed that in almost all cancers p53 gene is inactivated [4]. The main mechanisms are: Amino-acid-changing mutation in the DNA-binding domain; deletion of the carboxy-terminal domain; multiplication of the mdm2 gene in the genome; viral infection; deletion of the p14ARF gene; mislocalization of p53 to the cytoplasm, outside the nucleus [5].

A recent hope for a therapeutically solution appeared after Xue W et al. experiments published in the paper *Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas*, *Nature* (445: 656-660, 2007).

The paper is organized as follows: section 1 provides a brief overview of the main models from the literature pointing out their weak and strong points; section 2 deals with the improvements of the models biological as well as mathematical, including as a novelty a new type of activation function; finally, section 3 gives some direction for future research in the field.

1. Models from the literature

In this section, we will briefly present some models developed until now, with their strong points and weak points. The intention is to cover almost all was done in this field, in order for the reader to get a general idea about the challenging of p53 modeling.

From chronological point of view, the first model is assigned to the Bar-Or et al. (2000) [6]. The variables of the model are: designating P53, Mdm2 and I , where I is a hypothetical intermediary state of p53. The key finding was that under certain conditions p53 and mdm2 undergo damped oscillations. A weak point was the complexity of the model due to the relationships between the components. Totally there are 13 parameters that have to be estimated for only three components. The hypothetical intermediary I could be seen both as a weak and strong point; a weak point because its real existence cannot be justified and a strong point because it helped another authors to create a better models. This hypothetical intermediary can now be explained as a sum of a time lag and the inclusion of active p53 in the models that followed this first model. Another good point except the generation of oscillations was the introduction of two functions in the models, signal and threshold. The signal function in recent models are denoted by the ATM protein and the threshold function is explicitly taken as a Hill function or, not so common, as a step-wise function.

In the same period another model which came out with new features appeared. It is the model developed by Mihalas et al (2000) [7]. The strong points of the model are that it can show different type of behavior such as: rapid evolution towards steady states, damped oscillations and sustained oscillations. The model also introduced the time delay and has two types of activating function: continuous using a Hill function and discontinuous- threshold using step-wise function. The weak point of this model lies in the simple approach of the p53 network, it taken into account only P53 and Mdm2 proteins and their corresponding mRNA. Another weak point can be the numerical simulation using non-specific software written in Pascal combined with the lack of practical experiments.

Tiana et al.(2002) [8] developed a two equations model which came out with a new type of function for p53-mdm2 interaction and proved that the time delay is an

essential ingredient of the system that has a robust behavior. The weak points are: not mathematical demonstrations, the numerical simulation with non-specific software and not taking into account ATM protein.

One year later in 2003 Monk [9] created a model with three equations using as variables except P53 and Mdm2 proteins, the mdm2 mRNA. He also used a time delay describing the transcription process and combined this delay with a Hill function. He obtained a behavior according to experimental data but it was unclear from the model how the cell stress would affect the system; as the experiments suggest it is important that, at equilibrium, the oscillations do not occur and they are seen only a certain amounts of damage.

In 2004 Lahav et al. [10] made an experiment using MCF-7 cells in which suggests that the response of p53 network to DNA damage is similar to a digital clock behavior. In a few words, the increasing of the DNA damage will not increase the amplitude of pulses but their frequencies. This experiment triggered the development of many models in the next year.

Ciliberto et al. (2005) [11] proposed a model to explain this behavior. As variables, it took into account the DNA damage, 3 forms of p53 and 3 forms of Mdm2. The model addresses for the p53 the degradation part of the protein as two of variables are the mono- and poli-ubiquitinated form. Another strong point was the consideration of two compartments for the system: nucleus and cytoplasm. Further more, besides the classical negative feedback loop between p53 and mdm2, the model considered also a positive feedback loop involving PTEN and Akt. In this manner the model behavior was in accordance with the experiments from [10]. We have to mention that in some living cells and especially in MCF-7, PTEN is not induced by p53 and therefore the model is not quite appropriate as the experiments were performed on MCF-7 cells.

Ma et al. (2005) [12] proposed a more complex model who's behavior was also similar to experimental findings. For the first time it was introduced the concept of three modules for p53 network modeling. Those are: DNA damage module, ATM switch and p53 oscillator. For the first module it was used a Monte Carlo method for simulation of the stochastic process related to DSB generation. In the second module three types of ATM forms were taken into account and in the third module activated form of p53 was also considered as a distinct variable. The model in this form could not be a subject for a rigorous mathematical analyze.

Wagner et al. (2005) [13] based on the model from [9] and the concepts from [12] created a model which can exhibit a behavior similar with the results from [10] and also can support a fully mathematical approach. Nevertheless the model has only 4 variables, therefore is quite simple.

Based of the modular concept from [12] a more complex model was created by Chickarmane et al. [14]. In the first two modules besides DNA damage and ATM also Nbs-1, a repair protein, was considered. In the third it can be found 10 variables, the new ones are Arf, Siah, β -catenin and their interaction products. The model lacks the time delay but regarding its complexity can be seen as a forward step in p53 network modeling.

Geva-Zatorsky et al. (2006) [15] also performed a practical experiment using MCF-7 cells which showed undamped oscillations. They also proposed not a single mathematical model but six, trying to cover all the aspects of p53 network modeling including time delay and ATM. The models are simple but they are in accordance with experiments.

Recently, Bose et al.(2007) [16] proposed a model that take into account a parallel pathway that leads to arrest of G₂/M transition. The p53 network was studied from a biological perspective. The findings dealt with the apoptotic response of the cells and some experiments were suggested to be performed in order to prove the theoretical results.

2. Improvements of the models

All the models presented before can be improved; some of the improvements were already accomplished. The improvements can be split into two categories: biological improvements and mathematical improvements.

The biological improvements were achieved by adding new components of p53 network into the model - described for every model presented, explicitly introducing the time lag or, addressing the role of ATM-mediated p53 activation in initiating and regulating oscillations.

Regarding the time delay this was explicitly introduced in the models from [7], [8], [9], [13], [15] and [17]. The model from [7] was further improved by our team in [18] by introducing three delays corresponding to the transcription, translation and transport times. Other type of biological improvement of the models was taking into account the role of ATM. This was done in the models from [12], [13], [14], [15], [19].

Mathematical improvements were not so extended in the models presented here. We can mention the introduction of Poisson distribution for a better describing of DNA-damage module in [19], engineering approach using Laplace transforms in [17]. Other mathematical related improvements can be time-dependent noise in protein production rates found in [15] and the replacing of Monte Carlo method used in [12] for simulation of DSB production with the simulation in Matlab [19].

In this field our team has come with an important contribution. The mathematical improvements made by us rely on introduction of distributed time delay, delay kernel and step-wise function for activation. The step-wise function was introduced in [7] and we manage to use Maple for numerical simulation [20].

The introduction of the distributed time delay was performed for the model from [7] and the results can be found in [21]. The biological reasoning for this was that the processes described by the model are continuous, therefore an integral form for delay is more appropriate.

The next step in building a model with time delay was made by using the delay kernel. The classical delay found in other models is just a particular case of kernel delay (Dirac distribution). In this manner we can analyze the system behavior with other types of distributions such as normal distribution, Erlang and exponential. The results were presented in [22] and [23].

The mathematical improvements were not only limited to introduction of the new features in models but also consist in a full mathematical investigation which is materialized in a working procedure. The steps are: (1) establishing the mathematical model; (2) looking for the stationary states of the model from Step 1; (3) investigation using classical methods of the local stability of the positive stationary states by analyzing the corresponding transcendental characteristic equation of the linear system associated to the model from Step 1; (4) finding of the conditions for the delay kernel's parameter, for which there is a Hopf bifurcation; (5) determination of the Poincare normal form associated to the system using the central manifold; (6) analyze of the

limit cycle using the Lyapunov coefficient; (7) numerical simulation of the limit cycle in the neighborhood of the stationary state using Maple 11 with the discussion of the results. We have to emphasize that the numerical simulations performed used small programs written by the authors.

With this procedure was analyzed the model from [7] with the details in [24], model from [8] with the details in [25] and also all the other models which had as basic model the one from [7]. The details can be found in [21] and [23]. Numerical simulation using Maple 11 were performed also for the model from [13].

The model from [13] was improved with the change of Hill function with a step-wise one. The step-wise function is a more realistic one for biological processes as many of them obey the law “all or nothing” but creates difficulties because it is a discontinuous function. One solution to overcome this was the use of a Hill function which can approximate the step-wise function and more than that it is a continuous one. Until recently we also used a Hill function but now we manage to deal with a step-wise function.

Other improvement that was performed is the use of a signal function in order to describe the behavior of p53 network not only in a single cell but also in a population of cells. The details can be found in [26].

It is not least of importance to mention that all the numerical simulations performed with Maple 11 were in accordance with those presented by the authors of the above models and also with the practical experiments.

In the end of this section we propose a new type of activation function using a fuzzy function for the models from [26] in order to achieve a smoother approach of the biological process. In addition, for all these models, Lyapunov exponents can be analyzed taking into account the time delay. In this manner can be assessed the behavior along the orbits of the systems in order to see if the behavior is chaotic or not. The full-mathematical proven and the graphical representations will be done in future papers.

3. Discussions - Future prospects

The target of the p53 network modeling should be the development of a new model which will be able to reflect all the major aspects in this field. This new model should incorporate the three modules concept from [12], two compartments: nucleus and cytoplasm and mono- and poli-ubiquitinated form of p53 as in [11], should take into account other p53 components as E2F1 and Arf as presented in a doctoral thesis by Daniel Brewer and Siah, β -catenin and MRN repair complex as in [14]. Also other pathways as PTEN [11] and chk [16] should not be omitted and also the major target genes of p53 [5] should be included in order to address the p53 main action as cell-cycle arrest and apoptosis. Last but not least, a delay kernel should be taken into account as the biological processes are not instantaneous.

The model should be a robust one, having the ability to permit parameter variation in order to fit the wide range of behaviors encountered in human cells. This will lead to a very complex model that will be the base for new simulation software with the target oriented towards personalized health care.

References

- [1] M.Ljungman, D.P.Lane: Transcription-guardian the genome by sensing DNA damage, *Nature Reviews*, **4** (2004), 727–737.
- [2] K. W. Kohn, Y. Pommier: Molecular interaction map of p53 and Mdm2 logic elements, which control the Off-On switch of p53 in response to DNA damage, *Science Direct, Biochemical and Biophysical Research Communications*, **331** (2005), 816–827
- [3] M.E. Perry: Mdm2 in the response to radiation, *Mol. Cancer Res.*, **2** (2004), 9–19
- [4] D. Alarcon-Vargas, Z. Ronai: p53-Mdm2-the affair that never ends, *Carcinogenesis*, **23** (2002), no.4, 541–547.
- [5] B. Vogelstein B, D. Lane, A.J. Levine: Surfing the p53 network, *Nature*, **408** (2000).
- [6] R. Lev Bar-Or, R. Maya, L.A. Segel, U. Alon, A.J.Levine and M. Oren: Generation of oscillations by p53-Mdm2 feedback loop: A theoretical and experimental study, *PNAS*, **97**(2000), no.21,11250–11255.
- [7] Gh.I. Mihalas, Z. Simon, G. Balea and E. Popa: Possible oscillatory behaviour in p53-mdm2 interaction computer simulation, *J. of Biological Systems*, **8** (2000), nr. 1, 21–29.
- [8] G. Tiana, M.H.Jensen and K. Sneppen: Time delay as a key to apoptosis induction in the P53 network, *Eur. Phys. J.*, **29** (2002), 135–140
- [9] N.A. Monk: Oscillatory expression of Hes1, p53, and NF-kappaB driven by transcriptional time delays, *Curr. Biol.*, **13**, (16) (2003), 1409–1413.
- [10] G. Lahav, N. Rosenfeld, A. Sigal, N. Geva-Zatorsky, A.J. Levine, M.B.,Elowitz and U. Alon: Dynamics of the p53-Mdm2 feedback loop in individual cells, *Nat. Genet.*, **36** (2004), 147–150.
- [11] A. Ciliberto, B. Novak and J.J. Tyson: Steady States and Oscillations in the p53/Mdm2 Network, *Cell Cycle*, **4** (2005), Issue 3, 488–493
- [12] L.Ma, J.Wagner, J.J.Rice, W.Hu, A.J.Levine and G.A.Stolovitzky: A plausible model for the digital response of p53 to DNA damage, *PNAS*, **102** (2005), nr. 40, 14266–14271
- [13] J.Wagner, L.Ma, J.J.Rice, W.Hu, A.J.Levine and G.A.Stolovitzky: p53-Mdm2 loop controlled by a balance of its feedback strength and effective dampening using ATM and delayed feedback, *IEE Proc.-Syst. Biol.*, **152** (2005) No.3., 109–118.
- [14] V. Chickarmane, A. Nadim, A. Ray and H.M. Sauro: A P53 oscillator model of DNA break repair control, *arXiv:q-bio.MN/0510002v1*.
- [15] N. Geva-Zatorsky, N. Rosenfeld, S. Itzkovitz, R. Milo, A. Sigal, E. Dekel, T. Yarnitzky, Y. Liron, P. Polak, G. Lahav and U. Alon: Oscillations and variability in the p53 system, *Molecular System Biology* (2006), msb4100068-E1.
- [16] I. Bose, B. Ghosh: The p53-Mdm2 network: from oscillations to apoptosis, *J. Biosci.* **32(5)** (2007).
- [17] B. A. Ogunnaike: Elucidating the digital control mechanism for DNA damage repairs with the p53-Mdm2 system: single cell data analysis and ensemble modelling, *Journal of Royal Society*, (2005).
- [18] Gh. I. Mihalas, M. Neamtu, D. Opris and F.R. Horhat: A dynamic p53-Mdm2 Model with Time delay, *Chaos, Solitons and Fractals J.*, **30**, (2006), nr. 4, 936–945.
- [19] J.P.Qi, S.H.Shao, D.D.Li and G.P.Zhou: A dynamic model for p53 stress response networks under ion radiation, *Amino Acids* (2006).
- [20] F.R. Horhat, Gh.I. Mihalas: Modeling and simulation of some oscillatory biological processes, *29th International Conference of the Romanian Medical Informatics Society, Medinf2007*, 8-11 November 2007, Sibiu, Romania.
- [21] M. Neamtu, D. Opris and R. F. Horhat: A dynamic p53-mdm2 model with distributed time delay, *math.DS/0601591*, 24 Ian 2006.
- [22] F.R. Horhat, M. Neamtu and G. Mircea: The P53-Mdm2 network.Mathematical models and numerical simulations, *Differential Geometry-Dynamical Systems* October 5-7, 2007, Bucharest, Romania.
- [23] F.R. Horhat, G. Mircea, M. Neamtu and D. Opris: Hopf bifurcation for the dynamic p53-mdm2 interaction model with distributed delay, weak and strong kernel, *Conference Francophone sur la Modelisation Mathematique en Biologie et en Medicine*, 12-14 July 2006, Craiova, Romania.
- [24] D. Opris, M. Neamtu and F.R. Horhat: Hopf Bifurcation in p53-Mdm2 Interaction Model with Time Delay, *Proceedings of the 9th National Conference of the Romanian Mathematical Society*, 6-7 May 2005, Lugoj, Romania, 461–470.
- [25] M. Neamtu, R. F. Horhat and D. Opris: Hopf bifurcation analysis for a mathematical model of P53-Mdm2 interaction, *International Journal of Bifurcation and Chaos*, in press, January 2008.
- [26] R.F.Horhat, M Neamtu, D.Opris: p53 gene. Biological aspects. Mathematical models. Numerical simulation, *EdituraMirton*, Timisoara, 2008, ISBN 978-973-52-0301-6.

Using a Low-Cost Simulation Approach for Assessing the Impact of a Medication Administration System on Workflow

A. W. KUSHNIRUK ^{a,1}, E. M. BORYCKI ^a

S. KUWATA ^b, and H. WATANABE ^b

^a*School of Health Information Science, University of Victoria, Victoria, Canada*

^b*Tottori University Hospital, Tottori, 683-8504, Japan*

Abstract. This paper describes the analysis of the impact of a medication administration system on clinical workflow. The methodological framework employed was based on in-depth analysis of simulated user interactions with a medication administration system. The approach involved the collection of rich data consisting of audio and video recordings of interactions between 16 subjects (5 nurses and 11 physicians) as they interacted with a medication administration system. Methodological considerations and issues in conducting such studies are discussed. The study indicated that use of the system would have a significant impact on nurse and physician workflow and that this impact could be accurately identified using simulation approaches prior to widespread release of such systems in real clinical environments.

Keywords. Human interfaces, user interfaces, EPR-CPR-EMR

Introduction

Simulation methods have been used in biomedical informatics to study aspects of human computer interaction in a number of health research domains including human factors, usability engineering, doctor-patient interactions involving technology, health professional decision-making, and new device testing [1-5]. In this paper we describe application of a new approach towards evaluating the effects of health information systems upon clinical workflow. The system under study was a medication administration system that was integrated with bar-coding technology. The system allows the doctor or nurse to scan an RFID bracelet to identify the patient and to scan labels on medication bags to verify and document medication administration.

¹ Corresponding Author: Professor Andre Kushniruk, School of Health Information Science, University of Victoria, PO Box 3050 STN CSC, Victoria, British Columbia, V8W 3P5, Canada; E-mail: andrek@uvic.ca

1. Background and Motivation

Health information systems can have a significant impact upon clinician work. The impact of health information systems upon clinician work has been both positive (e.g. reducing medical error, length of stay) and negative (e.g. altering traditional workflow patterns that promote clinician communication and coordination of work). This has led a number of researchers to theorize about the socio-technical implications (i.e. task-technology fit) between health information systems and real world clinical environments [6]. More specifically, some researchers have suggested many of the difficulties experienced when implementing health information systems arise from a poor fit between clinical tasks and technology, leading to unintended clinical practice and workflow effects [6-7]. Over the past decade considerable researcher effort and energy has been directed towards documenting the unintended effects of technology upon clinician workflow using observational approaches in real life settings well after a health information system has been deployed. More recently, researchers have been advocating the use of simulations as a methodology for assessing the potential impact of health information systems upon clinician work [5]. Simulations allow clinicians and informaticians to determine the impact of a system and improve workflow changes *before* the system is deployed in a real world context. Specifically, we explore the value of using simulations that involve users interacting with information systems and devices in simulated clinical environments as they perform tasks in order to determine their impact before system release.

2. Methods

Subjects: Sixteen subjects participated in this study: 5 nurses and 11 doctors. All were employed at the hospital where the system under study was to be implemented.

Simulation Environment: The study was conducted in a real hospital room. Equipment included a laptop computer on a cart running the commercial medication administration system under study. The “patient” consisted of a dummy (i.e. a mannequin). In the room there was a medication cart with IV bags and medications.

Recording Equipment: To record the human-computer interaction in its entirety, the commercially available screen recording program Hypercam © was installed on the laptop on the cart running the medication order entry system. Hypercam© allows computer screens to be recorded digitally as movie. In addition, a microphone was placed on the cart to pick up subjects’ verbalizations while they interacted with the system and the patient. A Sony mini-DVD camcorder (mounted on a tripod) was used to record the physical activities of the subjects (e.g. nurses and doctors) as they alternated between working on the computer and interacting directly with the patient.

Task and Procedure: The subjects were instructed to log into the medication order entry system, scan the patient’s identification wrist band (to confirm their identity) and then carry out a range of medication orders indicated for that patient (involving administration of intravenous medications as well as interactions with the medication order entry system to obtain orders and confirm their completion). This included routine tasks such as administering a once a day oral medication to a patient to more

atypical tasks such as mixing a series of intravenous (IV) medications and then administering via an IV pump. At the end of completing the set of medication administration tasks (which were recorded in their entirety) a semi-structured interview was conducted with the subject. The prompts used in the interview included the following: (a) Did you have any difficulty with the task? (b) Did you have any difficulties with the barcode reader? (c) Did you have any difficulties during the work process?

Data Analysis: The audio portion of the simulation session containing verbalizations from the subject were transcribed in their entirety. The audio portion of the simulation were then uploaded to a research video annotation program known as Transana © which allowed for the audio portion of the session to be linked to both video views (i.e. of the computer screens as captured using Hypercam © and the physical view of the subjects captured using the camcorder). This allowed for both views of the simulation to be viewed, time stamped, coded and analyzed simultaneously (as shown in Figure 1).

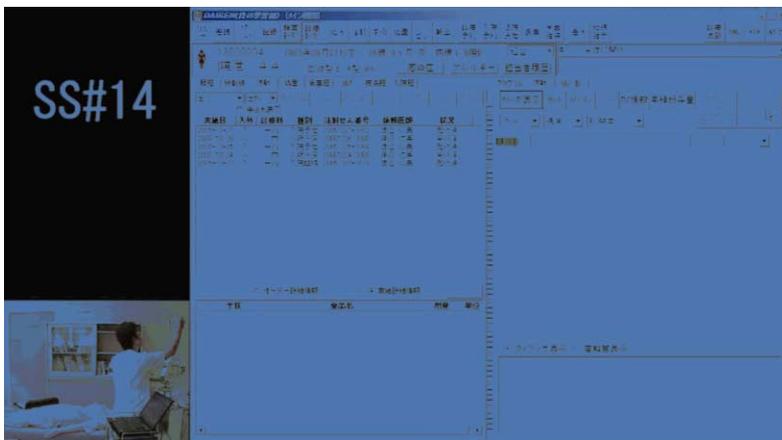


Figure 1. Integrated views of the computer screens and of the subject's physical interactions.

Applying a usability coding scheme, modified from Kushniruk and Patel [2], the researchers watched the video and annotated the corresponding transcripts and post-task interviews to identify: (a) human-computer interaction actions, such as selection of patients from a drop-down list and confirmation of orders, (b) physical activities, such as hanging an IV bag etc., (c) potential problems in the workflow, and (d) conceptual themes regarding impact of the system on workflow.

3. Results

Each hour of video data took approximately 2 hours for two experimenters to code and analyze. The coding was reviewed by the two experimenters with minor disagreement

being resolved during subsequent discussion. An excerpt from the coded transcript of a nurse while carrying out the required tasks is given below (along with the post task interview). The numbers in the example correspond to the actual time elapsed on the video counter. Annotations and codes are given in caps:

MEDICATION ORDER INFORMATION OBTAINED BY NURSE

00:14 NURSE SEARCHES FOR PATIENT ON THE COMPUTER

00:45 NURSE VIEWS ORDER LIST ON THE SCREEN

00:51 NURSE SELECTS MEDICATION ORDER FROM LIST

00:55 VERIFICATION SCREEN APPEARS

NURSE WALKS OVER TO PATIENT TO CHECK IDENTIFICATION

00:59 NURSE TALKS TO PATIENT - "Nice to meet you. I will now give you an IV drip"

01:09 NURSE SCANS PATIENT IDENTIFICATION (FROM PATIENT'S WRIST BAND)

01:10 VERIFICATION SCREEN AUTOMATICALLY UPDATES

NURSE WALKS BACK TO COMPUTER

01:25 NURSE VIEWS EXECUTION INFORMATION ON THE COMPUTER

NURSE WALKS OVER TO PATIENT AND SETS MEDICATION BAG

NURSE WALKS BACK TO COMPUTER

03:15 NURSE CONFIRMS ADMINISTRATION OF MEDICATION ON THE COMPUTER

POST-TASK INTERVIEW:

Experimenter: Did you have any difficulty with the task ?

Subject: I'm used to this operation, but sometimes it is hard to use the barcode reader when the barcode is not clearly printed.

Experimenter: Did you have any difficulties with the barcode reader?

Subject: In today's operation there were no problems. But in the real situation, sometimes the scanner doesn't respond to the barcode. Also, sometimes the cord of the scanner is too short to reach the patient.

Experimenter: Did you have any difficulty during the work process?

Subject: In general, I want a more simplified system for the verification process. The more patients there are, the more difficult the verification would become.

4. Observational Data and Emergent Conceptual Themes

This section will present the findings from our simulation from both the observational component of the study (i.e. from analysis of the video recordings of the subjects while carrying out tasks) as well as from thematic analysis of the post-task interview transcripts. From this analysis, we identified 6 potential issues described below.

4.1. Sequence of Workflow Activities

The analysis of recordings of the users' interactions and activities indicated that the system imposed a fixed and relatively rigid sequential order of activities (that could not easily be deviated from) in order to document the medication administration. As illustrated in the excerpt given above, the user of the system proceeds from viewing the order list, verifying the patient, selecting an order, verifying the medication, scanning the medication bag, administering the medication, confirming the administration and

then proceeds to another medication. In addition, the interview data indicated that the majority of subjects felt the verification process could be simplified.

4.2. Response During Urgent Situation

A recurrent theme that emerged from the analysis of the post-task interview transcripts indicated concern by subjects of how the system might affect workflow during periods of urgency where the physician or nurse might need to leave the prescribed sequence imposed by the computer system to deal with emerging clinical situations. In considering the ordering of tasks that the system imposed for entry of all medications, one nurse subject stated that “sometimes in the emergency we have to skip one of these procedures due to its time-taking process and someone might need urgent help, but with this system I don’t think I’d be able to do that”.

4.3. Increased Work Activity

In conducting the simulations, as the number and complexity of medications to be entered increased, as the workflow imposed by the system is sequential (without potential for parallel activities) there would appear to be an incremental increase in work activity in dealing with long lists of medications that might be detrimental to patient care (e.g. ten or more medications). One subject stated that “in the first part of the experiment, the patient has few records so the screen will show up in about 10 seconds, but the real patients with a long history of their admissions have lots of records so it takes a minute or more to open. It would be acceptable if we had only one patient, but in real case we have 10 or more patients so I think this process is really burdensome for us”.

4.4. System Response Time

It was noted by some subjects (during the interviews) that system response time might become a problem, with patients having long histories requiring longer for the information needed to appear on the screen. From the observational data it appeared this could be compounded by situations with multiple patients requiring complex combinations of medications. One subject stated that “I have to access many screens, which takes more time”.

4.5. System Lock

It was noted by one subject that during occasions when another health professional may be accessing the patient’s record that they would become locked out of that record (i.e. inability to access medication information when the patient’s record is completely “locked out” by other users of the system who are accessing the system at the same time).

4.6. Potential Need for System Override

Several subjects commented that under circumstance of increased work activity, complex medication administrations and emergency conditions (as described above)

that there might be the need for a way to override the lock-step sequencing of activities required by the system under emergency conditions.

5. Discussion

In this paper we have described our work in applying a simulation approach to evaluating the impact of a new health information system upon clinician workflow. This was based on analysis of user interactions involving a range of routine and atypical clinical tasks. The approach builds on previous work in the area of simulation [3] as well as usability testing [2] and led to collection of a rich set of observational and interview data. It should be noted that although the simulation could be considered “high fidelity” (in that it was highly realistic, being conducted within a *real* environment) the cost of running the study was minimal with recording equipment costing under \$1000 Canadian dollars and analysis costs under \$2000 Canadian dollars. The study found that in addition to identifying potential sources of specific problems that would arise from implementation of the new system, it was observed that introduction of the computer would likely lead to a major change in the process of medication administration. It should be noted that this was discovered *prior* to widespread implementation of the system and based on the results of this study the hospital informatics staff were able to fine-tune and customize the system to minimize the problems. Using realistic clinical simulations, such as the one described in this paper, to assess both intended and unintended consequences of introduction of information technology is urgently needed in healthcare in order to reduce possibility of error and increase both user acceptance and satisfaction with systems [8-9]. Knowing the impact of systems before they are released can also greatly inform system modifications and customization as well as preparing organizations about what to expect prior to widespread system implementation.

References

- [1] A.W. Kushniruk, D.R. Kaufman, V.L. Patel, Y. Levesque, P. Lottin, Assessment of a computerized patient record system: A cognitive approach to evaluating an emerging medical technology, *MD Computing* **13**(5) (1996), 406-415.
- [2] A.W. Kushniruk, V.L. Patel, Cognitive and usability engineering approaches to the evaluation of healthcare information system, *Journal of Biomedical Informatics*, **37** (2004), 56-76.
- [3] J.G. Anderson, Evaluation in health informatics: Computer simulation, *Computers in Biology and Medicine*, **32** (2002), 151-164.
- [4] J.G. Anderson, C.E. Aydin, S.J. Jay, Evaluating health care information systems, Sage Publications, London, 1994.
- [5] A.W. Kushniruk, M. Triola, E. Borycki, B. Stein, J. Kannry, Technology induced error and usability: The relationship between usability problems and prescription errors when using a handheld application, *International Journal of Medical Informatics*, **74**(7-8) (2005), 519-26.
- [6] M. Berg, Patient care information systems and health care work: A sociotechnical approach, *International Journal of Medical Informatics*, **55**(2) (1995), 87-101.
- [7] R. Koppel, et. al., Role of computerized physician order entry systems in facilitating medication errors, *JAMA*, **293**(10) (2005), 1197-203.
- [8] E. Borycki, A.W. Kushniruk, Identifying and preventing technology-induced error using simulations: Application of usability engineering techniques. *Healthcare Quarterly*, **8** (2005), 99-105.
- [9] A.W. Kushniruk, E. Borycki, Low-cost rapid usability engineering: designing and customizing usable healthcare information systems, *Healthcare Quarterly* **9**(4) (2006), 98-100.

Process Mining Techniques: an Application to Stroke Care

Ronny MANS^{a,1}, Helen SCHONENBERG^a, Giorgio LEONARDI^b, Silvia PANZARASA^b, Anna CAVALLINI^c, Silvana QUAGLINI^b, and Wil van der AALST^a

^a *Eindhoven University of Technology, The Netherlands*

^b *Dept. Computer and System science, University of Pavia, Italy*

^c *IRCCS Casimiro Mondino Foundation, Pavia, Italy*

Abstract. In a competitive health-care market, hospitals have to focus on ways to streamline their processes in order to deliver high quality care while at the same time reducing costs. To accomplish this goal, hospital managers need a thorough understanding of the actual processes. Diffusion of Information and Communication Technology tools within hospitals, such as electronic clinical charts, computerized guidelines and, more generally, decision support systems, make huge collections of data available, not only for data analysis, but also for process analysis. Process mining can be used to extract process related information (e.g., process models) from data, i.e., process mining describes a family of *a-posteriori* analysis techniques exploiting the information recorded in the event logs. This process information can be used to understand and redesign processes to become efficient high quality processes. In this paper, we apply process mining on two datasets for stroke patients and present the most interesting results. Above all, the paper demonstrates the applicability of process mining in the health-care domain.

Keywords. Data analysis-extraction tools, Process, Event-based systems

Introduction

Nowadays, Health-Care Organisations (HCOs) place strong emphasis on medical and organisational efficiency and effectiveness, to control their health-care performance and expenditures. The aim of the HCOs is to provide the highest quality services at the lowest cost [1]. Consequently, it is of the utmost importance to evaluate existing infrastructures, and the services offered by these organisations. Therefore, it is crucial to explore and process the data collected by HCO systems. These data can be process logs from a process management system, or databases from the electronic clinical chart system, or unstructured data. In fact, in modern day organisations, information and communication technologies are becoming pervasive and there is an immense growth of their use. Contemporary Information Systems (IS) have no existence of their own, but they act in the context of an organisation and its business processes [2]. Such

¹ Corresponding author: Ronny Mans, address: Paviljoen J3, 5600 MB, Eindhoven. E-mail: r.s.mans@tue.nl

systems, driven by process models to enact and manage operational business processes, are referred to as *Business Process Management Systems* (BPMSs) [3]. BPMSs record the data of the executed activities in form of event logs. An event log is like a history of what happened in the information system. This recorded data can be helpful to gain a clear picture of the underlying business process.

BPMS now also focus on Business Process Analysis (BPA), which covers functions of diagnosis and simulation of business processes. One of its emerging areas is Business Activity Monitoring (BAM), which typically focuses on performance issues without considering causal and dynamic dependencies in processes and organisations. This is where process mining techniques can be employed in order to extract process related knowledge (e.g. process models) from event logs [4]. Process mining has been applied in many domains, mainly the service industry, but also in the health-care domain [5,6,11]. In this paper, we *applied process mining to discover the procedures for treating stroke patients in different hospitals. Additionally, we analysed the patient related events from stroke onset till arrival in the hospital.*

The remainder of this paper is organised as follows. An overview of process mining is presented in Section 1. Section 2 presents the data used for our analyses. In Section 3 we present and discuss the obtained results. Conclusions and future work are presented in Section 4.

1. Process Mining and the PROM tool

Process mining addresses the problem that most organisations have limited information about what is actually happening. In practice, there is often a significant gap between what is prescribed or supposed to happen, and what *actually* happens. Only a concise assessment of reality, which process mining strives to deliver, can help in verifying process models, and ultimately be used in system or process redesign efforts. The goal of process mining is to extract process related information (e.g., process models) from process logs, i.e., process mining describes a family of *a-posteriori* analysis techniques exploiting the information recorded in the event logs. Typically, these approaches assume that it is possible to sequentially record events such that each event refers to an activity (i.e., a well defined step in the process) and is related to a particular case (i.e., a process instance). Furthermore, some mining techniques use additional information such as the performer or originator of the event (i.e., the person/resource executing or initiating the activity), the timestamp of the event, or data elements recorded with the event (e.g., the size of an order).

Process mining is applicable to a wide range of systems. These systems may be pure information systems (e.g., ERP systems) or systems where the hardware plays a more prominent role (e.g., embedded systems). The only requirement is that the system produces *event logs*, thus recording (parts of) the actual behaviour. Usually, hospitals have a wide variety of systems that record events. For example, in an intensive care unit, a system can record which examinations or treatments a patient undergoes and also it can record occurring complications for a patient.

Traditionally, process mining has been focusing on *discovery*, i.e., deriving process models, organisational context, and execution properties from enactment logs. It is important to mention that there is no a-priori model, but, based on an event log, some model, e.g. a Petri net, is constructed. However, process mining is not limited to process models (i.e., control flow) and recent process mining techniques are more and

more focusing on other perspectives, e.g., the organisational perspective, performance perspective or the data perspective. For example, there are approaches to extract social networks from event logs and analyze them using social network analysis [7]. This allows organisations to monitor how people, groups, or software/system components are working together. Also, there are approaches to visualize performance related information, e.g. approaches that graphically show the bottlenecks and all kinds of performance indicators, e.g., average/variance of the total flow time or the time spent between two activities.

To be able to understand whether the HCOs under study achieve their goals of providing timely and high quality medical services, we conducted several experiments using the Process Mining tool called ProM, extensively described in [8]. ProM is a platform independent open source framework which supports a wide variety of process mining and data mining techniques, and can be extended by adding new functionalities in the form of plug-ins. The ProM framework can read event logs that are in the MXML format. With the ProM *import tool*, data from various systems, such as MS Access, can be converted into the MXML format.

2. The data used for process mining

Two data sets are available. One refers to the clinical course of stroke patients from their hospital admission to discharge (clinical data set), and the other one refers to the pre-hospital phase (pre-hospital behaviour data set).

2.1. Data set 1: Clinical data set

Data on 368 consecutive subjects with a confirmed diagnosis of first-ever ischemic stroke have been collected using an electronic clinical chart, developed with the relational database management system MS-Access and shared by the neurological departments or the stroke units of four Italian hospitals. Data, that comprehend diagnostic and treatment procedures, complications, etc., are labelled according to the time elapsed from the symptoms onset: the first 6 hours are the *acute stroke phase* (110 patients arrived in the hospital in this phase), and the subsequent period, up to 7 days after stroke, was considered as the *sub-acute phase*. In the past, this data set has been analysed to investigate the relationships between compliance to clinical practice guidelines and stroke outcomes, in terms of both survival and cost [9,10]. However, those analyses have been carried out through classical statistical procedures, while *this paper shows that kind of new insights can be achieved using process mining techniques*. Although it is mostly a "classical" medical database, some process information can be derived, due to the availability of timestamps for diagnostic, therapeutic and clinical events, e.g. hospital admission and discharge. More specifically, these timestamps represent the actual dates of when the associated activities occurred.

2.2. Data set 2: Pre-hospital behaviour data set

In the pre-hospital behaviour data set, collected through direct interviews with 234 patients, we find all events which occur from stroke onset till the patient's arrival in the hospital. The data set contains temporal information about the actions taken by patients, their relatives and general practitioners (GP). More specifically, as detailed temporal

data exists for each event, this offers the opportunity for discovering a model showing several performance indicators, like bottlenecks or time spent in between events.

3. Results

Due to the lack of space, we only show the most significant results of our analysis that highlight the potential of process mining.

Clinical data set results - Process mining can be used to construct process models for a whole data set, or parts that are of particular interest. One result that we would like to highlight in this paper is obtained by partitioning the data set on different hospitals for patients that arrived in the acute phase. In Figure 1, we only show the treatment process for hospital 1 and 2. The models are obtained by using the Heuristics miner and only show the main flow (relationships between events) of the process and only for the frequently occurring events. Events are depicted by boxes. The numbers in the activity boxes indicate the occurrence frequency of the activity, e.g. admission occurs 31 times in the log of hospital 2. The upper number on the arcs indicates the reliability of the relation between the activities, e.g. for hospital 1 the reliability of admission followed by neuroprotection is 0.917. The lower number on the arcs represents the number of times this activity pattern occurred in the log. The reliability of a relationship (e.g. event *i* followed by event *j*) is not only influenced by the number of occurrences of this pattern in the log, but is also (negatively) determined by the number of occurrences of the opposite pattern (*j* followed by *i*).

By comparing the obtained process models we observed a difference in treatment strategies between different hospitals. Most notably, hospital 2 performs hypertension therapy earlier and much more than the other hospitals. It is known that antihypertensive treatment is a common practice, although not always justified by

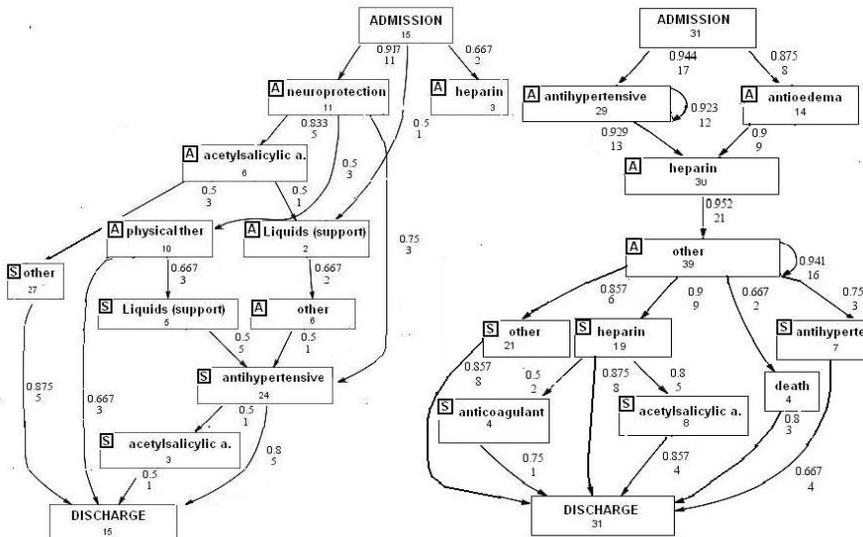


Figure 1. The treatment processes mined for the two hospitals H1(left) and H2(right). With respect to the original output, to facilitate the reader, labels "A" and "S" have been added indicating events belonging to the acute and subacute phase respectively.

different parameters can be run to see what the consequences are after removal of a bottleneck, e.g. change in throughput time.

4. Conclusion and future work

This work showed that process mining techniques can be applied successfully to clinical data to gain a better understanding of different clinical pathways adopted by different hospitals and for different groups of patients. It is interesting to analyse the differences, to establish whether they concern only the scheduling of the various tasks or also the tasks themselves. In this way, not only different practices may be discovered that are used to treat similar patients, but also unexpected behaviour may be highlighted.

Also, we have visualized the pre-hospitalisation pathways and identified bottlenecks. Even more interesting results could be found if additional data would be available from other health-care units involved in stroke management, like emergency rooms and rehabilitation clinics. But this requires high integration of the information systems involved, which is not (yet) the case. We believe that making health-care administrators aware about the potential of process mining can foster them to promote this kind of integration.

In this paper, we have applied process mining from a discovery point of view. An interesting future development would be to apply conformance testing. This would enable the comparison between (formal models of) medical guidelines and the execution in practice, i.e. the analysis of non-compliances.

References

- [1] Anyanwu K, Sheth A, Cardoso J, Miller J, Kochut K; Health-care Enterprise Process Development and Integration. *Journal of Research and Practice in Information Technology*, 35(2):83–98, 2003.
- [2] Hammori M, Herbst J, Kleiner N; Interactive Workflow Mining – Requirements, Concepts and Implementation. *Data and Knowledge Engineering*, 56(1):41-63, 2006.
- [3] van der Aalst WMP, ter Hofstede AHM, Weske M; Business Process Management: a Survey. In: *Proceedings of the 2003 International Conference on Business Process Management (BPM2003)*. Lecture Notes of Computer Science 2678, 1-12, Berlin, Springer-Verlag, 2003.
- [4] van der Aalst WMP, van Dongen BF, Herbst J, Maruster L, Schimm G, Weijters AJMM; Workflow Mining: A Survey of Issues and Approaches. *Data and Knowledge Engineering*, 47:2, 237-267, 2003.
- [5] Yanga WS, Hwang SY; A Process-Mining Framework for the Detection of Health-care Fraud and Abuse. *Expert Systems with Applications*, 31(1):56-68, 2006.
- [6] Lin F, Chou SC, Pan SM, Chen YM; Mining Time Dependency Patterns in Clinical Pathways. *International Journal of Medical Informatics*, 62(1):11-25, 2001.
- [7] van der Aalst WMP, Reijers HA, Song, M; Discovering Social Networks from Event Logs. *Computer Supported Cooperative Work* 14(6):549-593, 2005.
- [8] van der Aalst WMP, van Dongen BF, Günther CW, Mans RS., et al. ProM 4.0: Comprehensive Support for Real Process Analysis. In J. Kleijn and A. Yakovlev, editors, *Application and Theory of Petri Nets*, pages 484-494. Springer-Verlag, Berlin, 2007.
- [9] Micieli G, Cavallini A, Quaglino S; Guideline Application for Decision Making in Ischemic Stroke (GLADIS) Study Group. Guideline compliance improves stroke outcome: a preliminary study in 4 districts in the Italian region of Lombardia. *Stroke*, 33(5):1341-7, 2002.
- [10] Quaglino S, Cavallini A, Gerzeli S, Micieli G, GLADIS Study Group; Economic Benefit from Clinical Practice Guideline Compliance in Stroke Patient Management. *Health Policy*, 69(3):305-315, 2004.
- [11] Mans RS, Schonenberg MH, Song M, Aalst WMP van der, *Process Mining in Healthcare: A Case Study*. In: *Proceedings of Healthinf 2008, International Conference on Health Informatics, INSTICC*, 1(1):118-125, 2008.

Development of an E-learning System for Occupational Medicine: Usability Issues

M. Cristina MAZZOLENI^{a,1}, Carla ROGNONI^{b,c}, Enrico FINOZZI^a, Ines GIORGI^a,
Franco PUGLIESE^d, Marco PAGANI^c and Marcello IMBRIANI^a

^a *S. Maugeri Foundation IRCCS, Pavia, Italy*

^b *University of Pavia, Dept. of Computer Science and Systems, Pavia, Italy*

^c *CBIM – Consorzio di Bioingegneria e Informatica Medica, Pavia, Italy*

^d *Prevention and Protection Dept. – AUSL Piacenza*

Abstract. The aim of the present paper is to describe the process of developing an e-learning system for continuous medical education in the field of occupational medicine, with special focus on usability. The following steps are described: the needs analysis of the potential users; the prototype of the system that has been set up; the usability evaluation of the prototype by a sample of ten users; the analysis of the potential improvements; the evaluation of the revised system. The results of the usability tests point out that investing in improving usability was useful, even when they have not been recommended as mandatory. Only data collected from real active users will provide a more exhaustive evaluation, nevertheless it can be considered that positive results can be expected..

Keywords. e-learning, usability, education, occupational medicine

Introduction

As a consequence of the continuous technological and organizational changes of industrial production activities, occupational medicine ever increasingly needs a systematic and timely realignment of its scientific basis and intervention methods.

Moreover, in the last few years, occupational medicine physicians have been facing supplementary tasks, with respect to the clinical ones: they have the role of being consultant for risk assessment and management, information and education of workers, and the management of bureaucratic and legal issues. E-learning tools could support physicians to keep up to date with these dynamic changes. Continuous medical education (CME) through the internet has become popular [1] in some countries. Literature reports an interesting project [2] in the occupational medicine field, even though it was aimed at undergraduate students. In Italy, there is at present a certain drive in order to make e-learning for CME to take off. Even though a normative framework for distance CME has not been completely defined yet, some experimental

¹ Corresponding Author: M.Cristina Mazzoleni, Fondazione Maugeri IRCCS, Via Maugeri 4- 27100 Pavia Italy. cmazzoleni@fsm.it

projects are on-going with a twofold aim: the first, obviously, to support mandatory continuous medical education, and the second to try to overcome healthcare personnel skepticism as to e-learning.

When dealing with the introduction of computer based systems for healthcare personnel, user acceptance is often quoted as a critical issue, and e-learning systems are not an exception [3]. System usability and user acceptance are in direct causal relationship.

The aim of the present paper is to describe the process of developing and implementing an e-learning system for continuous medical education in the field of occupational medicine, with special focus on usability evaluation. The steps that have been undertaken in order to provide an e-learning tool that could be appreciated, and hence widely used, by the healthcare personnel are here described. In particular, the paper deals with: the needs analysis of the potential users; the description of the prototype of the system that has been set up; the usability evaluation of the prototype; the analysis of the identified criticisms and the evaluation of the revised system.

1. Methods

1.1. Needs analysis and educational model identification

ISO (International Standards Organization) defines usability as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use”.

From this definition, the identification of the needs of the target user group represents the mandatory starting point of the project itself, in terms of both topics to be treated as well as goals to be achieved. For this purpose a focus group, composed of experts in the field of occupational safety and health, has been created. The discussion inside the focus group led to the definition of a 15 item questionnaire, containing the topics related to both risk factors (mechanical vibrations, noise, carcinogenic agents, etc.) and specific aspects of the routine activity of an occupational physician (job suitability, communication, legal issues, etc.).

The questionnaire was distributed to 300 potential users to obtain the necessary information about the perceived educational needs and preferences. The analysis of the answers provided the identification of a subset of risk factors to be dealt with by the e-learning system. At the same time it revealed, as expected since the target group is composed of adults, a specific interest for real problem solving oriented issues, more than issues of acquisition of knowledge per se.

Consequently, the educational model has been designed in order to provide:

- Information transfer via self-learning modules for free navigation.
- Schematic representations of guide lines based behavioural paths, through hyper-flowcharts to improve knowledge transfer into clinical practice [4].
- Ad hoc case studies. They are particularly effective [5] in order to reinforce the learning process. Here they are aimed at testing the knowledge and the acquired ability in solving specific problems addressed by the above mentioned behavioural paths.
- Evaluation of the results of the learning process through traditional tests (questions with multiple choice answer).

Excluding the possibility of direct face to face interaction between teacher and learner,

but considering communication as a must in self-learning models, the figure of a tutor has been introduced into the educational model. He or she has the specific tasks of supporting the continuity of the learning process, and of giving further explanations and details according to the learner's requests.

1.2. The system

Titolo corso	Contenuti	Creatore	Durata	Crediti ECM
L'ASMA PROFESSIONALE	Contenuti	Massimo Corradi	5 h	5
POLMONITI da ipersensibilità e PNEUMOPATIE da metalli duri	Contenuti	Massimo Corradi	5 h	0
Le patologie correlate al RUMORE	Contenuti	Elio Valoti	5 h	0

Figure 1. A snapshot of the e-learning system

The e-learning system [6] has been developed using the course management system Moodle (Modular Object-Oriented Dynamic Learning Environment) [7] that is an open source package, designed using sound pedagogical principles, to produce internet-based courses and websites. Moodle uses SCORM (Sharable Content Object Reference Model) compliant modules [8].

The course list page shown in Figure 1 provides general information about the available courses with information and indications about the use of the system. After the registration to the website, users can enrol themselves into the proposed courses that are composed of different modules: tests, information modules, schematic representations of guide lines and ad hoc case studies. While test functionalities are embedded in Moodle, the other SCORM compliant modules have been implemented using different e-learning courses authoring tools, in particular WBTEExpress for information modules, Saba® Publisher for case studies and Microsoft Visio for guide lines representation. More information is provided in [9].

1.3. Usability testing

Usability has been evaluated in two phases: analysis of ease of use and intelligibility of basic functionalities (BF) and inspection of the subjective usability perception (SUP).

As far as BF is concerned, 10 users (occupational physicians, mean age 45, range 29 - 58) were enrolled for this evaluation. Two issues were assessed: intelligibility of provided information, and the ease to complete the tasks. The test was performed by the users in absence of any previous information about the system. A usage test was structured in order to:

- Verify that the users have understood: role of the pre-test (IT), how to earn CME credits (IE) and get general courses information (IG); the effects of web site registration (IR), relationship between web site and course registration

(II).

- Check the accomplishment of the tasks as to: course registration (TR), visualization and printing of one hyper-flowchart (TS), interaction with the tutor (TM), visualization of the educational offer (TC) and of the course content (TI). Only one attempt was allowed.

During the usage test session, each user, , filled in a specifically prepared form in order to verify the completion of the tasks and the correctness of the information he or she had understood. In the meantime, an experienced observer recorded the comments of the user as well as personal observations.

As to SUP, the Software Usability Measurement Inventory (SUMI) [10] has been used. It is a consistent tool to assess the quality of use of a software product from the user's point of view. It consists of 50 statements that the user has to answer. SUMI contains five sub-scales: *Efficiency* - support for the user to enable the user to get their work done; *Affect* - likeability, stress-free usage of the product; *Helpfulness* - degree of information about the product in the product itself; *Control* - amount of transparency as perceived by the end user; *Learnability* - ease with which a user can pick up how to use the software; and a *Global* scale. The raw scores obtained from the users are compared to the appropriate normative tables by SUMISCO software. In general, if some of the sub-scales are equal to or below 50 then they are poor in usability in that aspect. Sub-scales equal to or below 40 indicate the need for remedial action. Good software will achieve scores of 60 or more in most of the sub-scales.

At the end of the previously described BF usage test, each user has been asked to fill in the SUMI questionnaire.

2. Results

Figure 2 shows the rate of success for each dimension of the BF usage test,. Figure 3 shows the SUMI scores results. All of the SUMI sub-scales, except "helpfulness - degree of information about the product in the product itself", had a mean value equal to or greater than 50, and hence, according to SUMISCO, remedial actions were not strictly necessary. On the other hand, the BF test results showed some heavy criticisms, particularly as to TS, TM, IG, and II.

A further revision effort has been planned based on both the results of the BF test and of the expert recorded observations, therefore four kinds of remedial actions have been identified:

- Lexical related actions (L) – for example: to change the word "indice (contents)" into "contenuti (topics)" when referring to the list of contents.
- Layout related actions (S) – for example to change the column order in the table containing the course list.
- Procedure related actions (P) – for example to let the learner get in touch with the tutor during course navigation, instead of from the course list.
- Poor information related actions (I) – for example add specific information to improve the explanation of images and symbols.

In total 34 – respectively 13 (L), 4 (S), 10 (P) and 7 (I) – interventions have been identified as potentially useful.

The BF and SUP tests were performed again, using the same SUP sample of users, on the revised system in order to quantify the improvement.

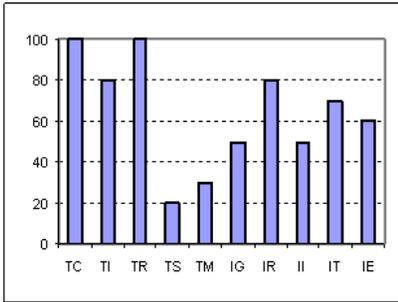


Figure 2. Initial success percentages

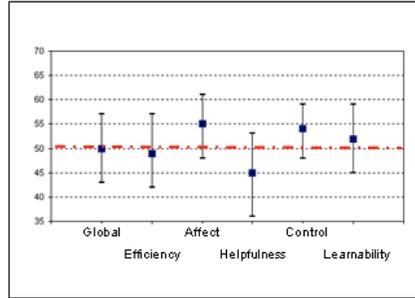


Figure 3. Initial SUMI scores

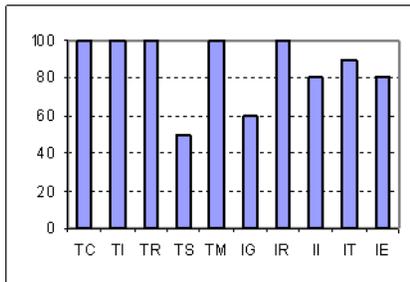


Figure 4. Final success percentages

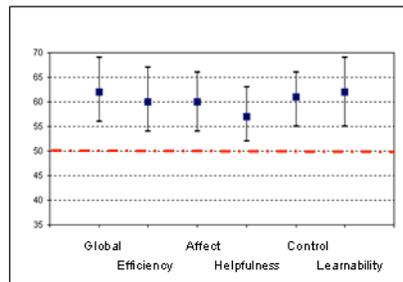


Figure 5. Final SUMI scores

Figure 4 and 5 above show the final results: SUMI scores have reached values typical of “good software”. An improvement of the system is documented also by BF test results, even though some of the percentages of success are lower than 100%.

3. Discussion and conclusions

The evaluation presented here shows that the efforts aimed at improving usability of the prototype software has led to good results. However further remarks are necessary. First of all the dimension of the sample: a sample of ten users is reported as “adequate” for usability studies [10, 11], but a larger group of test users would be desirable to attain more comprehensive results and to assess differences related to age, familiarity with computers, and so on.

Second: within the BF second test, rates of success for IG and TS were still quite low: for both cases it is suspected that this is due to misinterpretation of the question rather than to a problem of the system itself. In the case of TS for example, the task was “Open and print the schema of the diagnostic pathway for occupational asthma”. Many of the users didn’t accomplish the task at the first attempt because they used the “print” icon of the browser instead of the proper “print the schema” button available within the page. This occurred again even after the button had been moved to a more visible position, and the description had been changed to a more intelligible one. In a real situation, when a user opens the schema, probably he/she spends some time looking at the page content, and hence has the possibility to realize that there is a specific print button. The right task description should have been “Open the schema,

answer a question related to the content, and print the schema”.

Third: Only the usability of the basic functionalities of the system have been evaluated. An evaluation on a wider scale is advisable, but this study was primarily focused on preventing the user from abandoning the system at the very first phase because of unexpected difficulties or uncertainties. Similarly, the e-learning platform “back office” functionalities have not been taken into consideration in this study.

Fourth: the SUMI questionnaire has given good results and has proved to be a tool to quickly inspect subjective usability perception, but many users have had uncertainties when answering some items that did not seem to be pertinent. The median score values are comparable with the ones reported in other studies [12].

Finally: Comments, even by initially very skeptic users, have pointed out a positive attitude to use the e-learning system.

In conclusion, investing in improvements of usability has proved to be useful, even when not recommended as mandatory. In the present paper we have described the process we have followed to overcome one kind of obstacle for the use of e-learning systems, although this one is probably the most critical. The e-learning system is intended for a very wide variety of people, not known in advance, where the first impression do make the difference between “using or not using” the e-learning tool. Only data collected from real users will allow a more exhaustive evaluation, nevertheless positive expectations seem to be reasonable.

4. Acknowledgement

Research was supported by “E-learning per la qualità della vita e la sicurezza in ambito lavorativo” project of Ministero della Salute and S. Maugeri Foundation, IRCCS.

References

- [1] C.A. Olson, M.B. Shershneva, Setting quality standards for web-based continuing medical education, *J Contin Educ Health Prof*, Spring **24** (2), 100-111, 2004.
- [2] K. Radon, S. Kolb, J. Reichert, T. Baumeister, R. Fuchs, I. Hege, G. Praml, M. Fischer, D. Nowak, Case-based e-learning in occupational medicine-The NetWoRM Project in Germany, *Ann Agric Environ Med*, **13** (1), 93-8, 2006.
- [3] S. Childs, E. Blenkinsopp, A. Hall, G. Walton, Effective e-learning for health professionals and students--barriers and their solutions. A systematic review of the literature-findings from the HeXL project, *Health Info Libr J*, Dec, **22** Suppl 2, 20-32, 2005.
- [4] H.C. Vollmar, C.C. Schürer-Maly, J. Frahne, M. Lelgemann, M. Butzlaff, An E-learning Platform for Guideline Implementation – Evidence- and Case-based Knowledge Translation via the Internet, *Methods of Information in Medicine* **45** (4), 389-396, 2006.
- [5] A.S. Wilson, J.E. Goodall, G. Ambrosini, D.M. Carruthers, H. Chan, S.G. Ong, C. Gordon, S.P. Young, Development of an interactive learning tool for teaching rheumatology-a simulated clinical case studies program, *Rheumatology (Oxford)*, Sep **45** (9), 1158-61, 2006.
- [6] <http://fad.cbim.it>, last accessed at 2008-02-27.
- [7] www.moodle.org, last accessed at 2008-02-27.
- [8] www.adlnet.gov/scorm, last accessed at 2008-02-27.
- [9] C. Rognoni, G. Taino, E. Valoti, E. Capodaglio, M. Panigazzi, S. Ghittori, M.C. Mazzoleni, E-learning System for CME in Occupational Medicine, *MIE2008 Proceedings (in press)*, 2008.
- [10] www.ucc.ie/hfrg/questionnaires/sumi/index.html, last accessed at 2008-02-27.
- [11] J. Nielsen, Why you only need to test with 5 users, Alertbox, March 19, 2000. www.useit.com
- [12] H. Farrimond, T.L. Dornan, A. Cockcroft, L.E. Rhodes, Development and evaluation of an e-learning package for teaching skin examination, Action research, *Br J Dermatol*, Sep **155** (3), 592, 2006.

Economic Advantage of Pharmacogenomics - Clinical Trials with Genetic Information

Wataru OHASHI^{a,b,1}, Hiroshi MIZUSHIMA^a, Hiroshi TANAKA^a

^a *Tokyo Medical and Dental University, Information Center for Medical Science*

^b *Tokyo Medical and Dental University, Department of Bioinformatics, Graduate School of Medicine and Dentistry*

Abstract. The purpose of this study is to clarify the benefit and loss for the pharmaceutical companies when they adopt introducing pharmacogenomics in their clinical trials (in the following description, clinical trials by using pharmacogenomics is called "pgx clinical trial"), that is, when they use genetic information in their clinical trials. Particularly, the benefit for the pharmaceutical companies in terms of following two points is analyzed. 1. Development cost of new drug and period of clinical trial can be reduced because a clinical trial needs less subjects, 2. The new drug can be placed on the market earlier because the development period can be shortened. A survey conducted by Japan Pharmaceutical Manufacturers Association revealed that the pharmaceutical companies in Japan are interested in "pgx clinical trial". Specifically, 95% of the member companies (n=19) of the Association replied that the establishment of a guideline for pgx clinical trial by regulatory authorities are highly desirable. However, 65% of them (n=13) also replied that pgx clinical trial is difficult for the time being. It can be concluded that the pharmaceutical companies are positive about pgx clinical trial, but they cannot take a step towards it for several reasons: some of them may be worried their sales for non-responders will be reduced, poor understanding of pgx among the concerned parties, and not matured methodology of pgx clinical trial. This study shows that the advantage of pgx clinical trial outweighs its disadvantage. The sales may decrease because the drug is not used for non-responders, however, the number of subjects necessary for a clinical trial can be reduced, study period can be shortened and the drug can be marketed earlier. Furthermore, adverse events (AE) and adverse drug reactions (ADR) during the clinical trial and post-marketing phase can be markedly reduced. This represents a great benefit for the patients, pharmaceutical companies and the society as a whole.

Keywords. Biostatistics, Clinical trials, Mathematical models in medicine

¹ Corresponding Author: Wataru Ohashi, Tokyo Medical and Dental University, Department of Bioinformatics, Graduate School of Medicine and Dentistry, 1-5-45 Yushima, Bunkyo-ku Tokyo, Japan; E-mail: wohashi@bioinfo.tmd.ac.jp

Introduction

In the current system for drug approval, a new drug is approved when the response rate of the drug is significantly higher than that of existing drug or placebo. In a clinical trial during development of a new drug, the drug is approved and allowed to be marketed if it can successfully show statistically significant effect compared to a placebo or a control drug (existing treatment). In the current method of a clinical trial, the drugs are compared between "groups" of subjects, that is, between a new drug group and a control group. The difference between responders and non-responders is often explained in terms of difference in subjects' profile. When the subjects have largely similar profiles, the difference is explained by a term called "constitution".

On the other hand, the importance of "personalized medicine" is emphasized where a suitable treatment is selected according to the "constitution" of each patient. In the near future, it is expected that the drug and dose can be determined based on the genetic information including SNPs (single nucleotide polymorphisms) of each patient. According to the survey conducted by Japan Pharmaceutical Manufacturers Association in its member companies, 14 companies (70%) were afraid of reduced market size because the drug cannot be used for non-responders when genetic information is taken into account. 13 companies (65%) replied that pgx clinical trial is difficult because IRB (Institutional Review Board) does not approve pgx clinical trial (n=6, 30%), Ministry of Health, Labour and Welfare has not established a guideline yet (n=11, 55%), and pgx is not fully understood by patients and other relevant parties (n=17, 85%). However, given that 95% of the companies (n=19) replied that the establishment of a guideline for pgx clinical trial by regulatory authorities are highly desirable, the companies are not negative about pgx clinical trials itself [1]. It can be inferred that the pharmaceutical companies are interested in pgx clinical trial, but they cannot take action for several reasons: profit may decrease, pgx is not fully understood by concerned parties, and methodology of pgx is not mature.

The purpose of this study is to clarify the benefit and loss for the pharmaceutical companies when they adopt "personalized medicine", that is, when they take advantage of genetic information in their clinical trials. Particularly, the benefit for the pharmaceutical companies in terms of following two points will be analyzed. 1. Development cost of new drug and period of clinical trial can be reduced because a clinical trial needs less subjects, 2. The new drug can be placed on the market earlier because the development period can be shortened. Further purpose of this study is to appeal economic effect of the pgx to the pharmaceutical companies, and to advocate importance of the pgx and necessity of official guideline for pgx clinical trial to the regulatory authorities.

1. Examples of Personalized Medicine

For example, Herceptin is an anticancer drug by Roche used for prevention of postoperative recurrence or metastasis of advanced breast cancer. The drug is prescribed only for patients showing abnormal amplification of HER 2 gene (3+, highly positive), (2+, moderately positive). And after 90th, the study of the relations between drug metabolizing enzyme and drug metabolic rate has progressed. This study clearly shows the clear existence of Extensive Metabolizer (EM) and Poor Metabolizer (PM) which decided the differences of SNP of drug metabolizing enzyme. In case of

CYP2C29, PM rate is only 3% in white, but 20% in yellow race. If we can judge the drug metabolic rate by using the differences of SNP’s information between the races, it will lead the globalization of the clinical trials and be able to reduce the trials held in each country. In this study, the design of clinical trial is examined in which the patients

Table 1. Comparison of clinical trials.

Items	Current	Personalized
Determination of usage and dose	Effect and ADR are determined based on the average.	Can be optimized for each patient according to the patient's capacity of drug metabolizing enzyme (e.g., CYP2C29, CYP2D6)
Determination of the number and profile of subjects to be enrolled	Determined using sample size estimation equation	Can be selected and response rate can be improved based on relevant genetic information
Medical expenses	Rises due to unnecessary administration to non-responders	Can be cut because unnecessary administration to non-responder can be avoided (pharmaceutical companies may lose their sales for non-responders)
Risk of Adverse Events	Unknown adverse event	Can be minimized based on information on causative gene and minimization of drug doze.
Cost of Clinical Trial	Usual recruiting cost	Temporarily increased because of the Genetic test
Difficulties of Recruiting the subject	Have to get consent from test subjects	Have to get the other consent for genomic information.

are screened according to genetic information, and thus decreasing the total number of subjects. Table 1 shows comparison between current clinical trials and personalized clinical trials.

2. Methods

2.1. Principle of sample size estimation in conventional clinical trial

Generally, the sample size of a clinical trial is calculated according to an equation as follows, for example. [1][2][3]

$$n = \frac{\left\{ Z_{\alpha} \sqrt{2\bar{P}(1-\bar{P})} + Z_{\beta} \sqrt{P_1(1-P_1) + P_2(1-P_2)} \right\}^2}{(P_1 - P_2)^2} \quad (\text{where } \bar{P} = \frac{P_1 + P_2}{2})$$

(In case of chi-square test for the difference in ratio between two groups)

N: Number of necessary samples

P₁: Response rate in investigational drug group

P₂: Response rate in control drug group

Z_α: Value calculated from significance level (typically 1.96 at 5%)

Z_β: Value calculated from power (typically 0.84 at 80%)

For example, when the response rate in the investigational drug group is estimated to be 70% and that in the control group is estimated to be 60%, the sample size of each group is $n=356$ according to above equation. Similarly, when the response rate in the investigational drug group is 90% and that in the control group is 60%, then $n=32$. That is, if there is a large difference in response rates, thus small sample size will be sufficient. The size is calculated based on significance level (typically $\alpha=0.05$) and power (typically $\beta=0.80$) specified in the protocol, and estimated drug response rate and difference between the groups. The calculation differs depending on the test method used.

2.2. Principle of pgx clinical trial using genetic information

Figure 1. shows an example of response rate and necessary sample size. By extracting and excluding non-responders based on SNPs information, the response rate of the investigational drug can be raised by 10%, e.g., from 70% to 80%. Then, the sample size of each group can be cut by 274 (356 to 82). In total, this means 548 cases will be unnecessary. When the rate is raised from 80% to 90%, $82-32=50$ cases can be reduced in both groups. This means the total of 100 cases will be unnecessary. This effect is more apparent when the response rate is low. Thus, this approach is highly useful in clinical trials in which only little difference in response rate is estimated between groups such as clinical trials of psychotropic drugs. In a pgx clinical trial, additional cost is necessary for genetic testing before the clinical trial. Therefore, the total cost will be initially higher. However, if the sample size can be reduced by 548 in total, the cost and the study period as a whole will be less than the trial according to the existing method. Furthermore, the cost of genetic testing is decreasing and the banking and database of genetic information are expanding day by day.

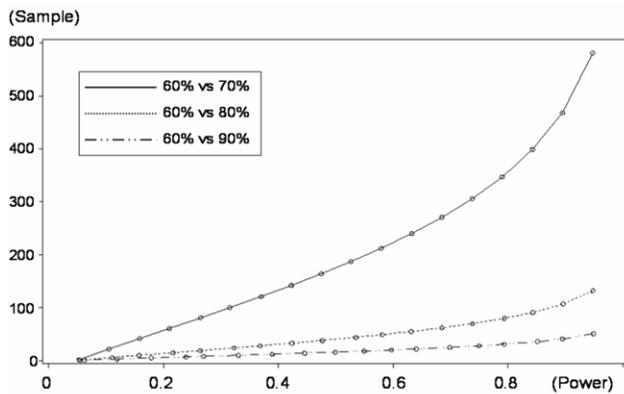


Figure 1. Response rate and necessary sample size (Control vs Investigational)

2.3. Reduction of risk of adverse events

When any known adverse event (AEs) is anticipated due to causative gene, the patients concerned can be excluded from the investigational drug group. Thus, the risk of adverse events can be reduced.

Assume that a clinical test of a drug of which response rate is estimated to be 60% to 70% is planned for Japanese. When CYP2C19 is involved in the drug metabolism, 20% of the 712 subjects are estimated to be poor metabolizers (PM), so that the number of PM in the study group is $712 \times 0.2=142$. If ADR appears 10% of them, 14 subjects

will suffer from ADR, and if it appears 20% of them, 28 subjects will be affected. In a pgx clinical trial, PM can be screened out in advance, thus ADR can be avoided at the stage of clinical trial. If the screening cost is not covered by health insurance and the cost is 40 thousand yen (350 dollars) per a subject [3], the extra cost for screening is 40 thousand yen \times 712 = 28.48 million yen (about 240 thousand dollars). However, should ADR occur, compensation about 120 million yen (1.1 million dollars) per a subject would be necessary in the case of death, and compensation about 490 million yen (4.3 million dollars) would be necessary in the case of permanent disability [4]. The extra cost of screening is not high compared to these expenses. When screening is conducted at Phase II (Ph2), ADR in Phase III and post marketing phase can be markedly reduced, which can compensate for the extra money used at the earlier stage. For the pharmaceutical companies, ADR can be avoided, thus, the compensation for the cost of hospitalization and treatment can be reduced. For the society as a whole, medical cost can be markedly saved. Cost of ADR is far from negligible. In the United States, more than 2 million patients have to be hospitalized due to ADR a year. 100 thousands deaths occur among them. The cost incurred by ADR exceeds 20 trillion yen, and 18 trillion yen of it is used for hospitalization (174.3 billion dollars). Medical cost for treatment of ADR amounts to 70 billion dollars a year [5].

3. Conclusion

As shown in Figure 2, the profit-cost structure of clinical trial is much improved in a pgx clinical trial compared to a conventional clinical trial. According to our estimation, a clinical trial that takes advantage of genetic information is superior to a conventional clinical trial in terms of cost and profit. In the above example, the development cost can be reduced directly by 1.6 million yen²/subject \times 548 subjects = 87.7 billion at the clinical trial stage. Furthermore, the development period can be shortened due to less subjects. The days necessary for one subject are 3 to 4 days in average [6]. Thus, if median 3.5 days is used for calculation, $584/3.5= 157$, that is, total 157 days can be reduced. The management cost and monitor dispatch fee during 157 days is hard to calculate, but these costs are reduced as well, further contributing to the reduction in the development cost [7].

On the other hand, the extra cost of screening and loss in profit due to smaller market can be calculated as follows. Cost of screening is 40 thousand yen \times 712 subjects = 28.48 million yen (about 240 thousand dollars). Days necessary for the screening can be estimated to be 10 days³. Thus, the days actually saved is $157 - 10=147$ days. Besides, the entire sales for non-responders is lost, which represents a net loss. However, the drug can be placed on the market 147days earlier than the conventional method because less subjects are used in a trial. Thus, the sale expected in these 147 days represents a net profit for the pharmaceutical company. Consequently, it can be concluded that the profit expected in the pgx clinical trial outweighs the extra cost and loss of the pgx clinical trial. Although Lipitor with the

² Excerpts from the material of "6th symposium of Kitasato University-Harvard School of Public Health" held on October 25, 2005. Mean cost per a sample in 29 clinical trials conducted in Japan by 17 member companies of Japan Pharmaceutical Manufacturers Association from April 2004 to March 2005. The cost includes fee for subjects, dispatch fee of monitors and management cost.

³ When the gene testing is contracted out to other company specialized in the test, 3 to 10 days are necessary for the test regardless of the sample size.

sales of about 1.5 trillion yen (13.7 billion U.S. dollars) is an exception, the sales of 18 drugs exceed 100 million yen (0.9 million dollars) a day [8]. Furthermore, the screening of relevant SNP can markedly reduce the AE and ADR during a clinical trial. This is a great benefit for the pharmaceutical companies given that the compensation for ADR is huge (hospitalization, death), and accompanying loss of credit is heavy for the company. What is more, the medical cost of the society as a whole can be saved as well. Thus, it can be concluded that ppx clinical trial is highly beneficial for the patient, company and the society.

For medical informatics, ppx clinical trials present a special challenge. Without appropriate security system for handling of genetic data, patients may be reluctant to participate in trials in fear of leakage of their data. Medical informatics may assume responsibility for establishment of security system crucial for ppx clinical trials.

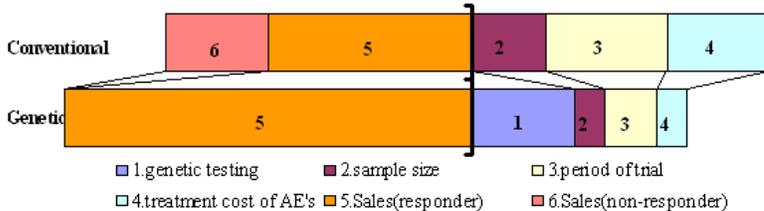


Figure 2. Profit-cost structure in conventional clinical trial and clinical trial using genetic information

In a clinical trial using genetic information, temporary increase in process and cost may be inevitable. But, when genetic testing comes to be widely used, the cost will drop. Additionally, as the databases of genetic information are growing, it is apparent that the genetic testing will be simplified. We expect the regulatory authorities to appreciate the advantage of ppx clinical trial, to establish well-defined guideline, and to conduct enlightenment activity for the people. Also, we expect the pharmaceutical companies to actively facilitate ppx clinical trials.

References

- [1] Questionnaire about pharmacogenomics, Japan Pharmaceutical Manufacturers Association, 2007. <https://www1.meteo-intergate.com/news/letter/119/>
- [2] *Statistical tables for the design of Clinical Trials*, M.D.Campbell, Blackwell Scientific Publications, Oxford, 1987
- [3] Approves Updated: Warfarin (Coumadin) Prescribing Information, FDA, 2007. <http://www.fda.gov/bbs/topics/NEWS/2007/NEW01684.html>
- [4] D.S.Dickinson, W.C.Bailey, B.I.Hirschowitz, Soong S-J Eidus L, M.M.Hodgkin, Risk factors for isoniazid (INH)-induced liver dysfunction, *Journal of Clinical Gastroenterology* **3** (1981), 271-279.
- [5] D.A.Hughes, F.J.Vilar, C.C.Ward, et al., Cost-effectiveness analysis of HLA-B*5701, genotyping in preventing abacavir hypersensitivity, *Pharmacogenetics* **14** (2004), 1-8.
- [6] The calculation is based on the data in "Commission to facilitate marketing of effective and safe medicine", MHLW, October 30, 2006. <http://www.mhlw.go.jp/shingi/2006/10/txt/s1030-5.txt>
- [7] J.A.DiMasi, R.W.Hansen, H.G.J.Grabowski, The price of innovation: new estimates of drug development costs, *Journal of Health Economics* **22** (2003), 151-185.
- [8] Y.Yatsuda, Sales rank of world medicines, Utobain News release, Jul 2007. <http://www.utobrain.co.jp/news-release/2007/070700/NewsRelease0707.pdf>

Recognising e-Health as Part of a Cohesive Professional Community

Jean M ROBERTS

Lancashire School of Health and Post-Graduate Medicine, University of Central Lancashire, Preston, UK

Abstract. This paper identifies a mechanism for specific professional registration in order to sustain a holistic community fit to practice in informatics to support the health domain. It considers risks and opportunities that have an international resonance, and comments on the areas where multi-national activity could provide additional impetus to improvement of the quality of the profession of health informatics overall. It puts the case for the over-arching term *health informatics* to be used to maximize synergy and drive up quality, whilst still recognising the specificity and place for more focused descriptive terms. Whilst grounded in the contemporary UK environment, the principles explored in this paper are currently being considered for adoption internationally.

Keywords: Managing change, professionalism, collaboration, capacity, competencies, maturity.

Introduction

From a low base in the 1960s, operational and academic ‘computing for health’ have developed in many ways across management and clinical disciplines, now commonly grouped as ‘informatics’. The informatics domain encompasses technology, information management, application functionality/usage, information governance and information quality. Many professional staff, clinical and management, have involvement in information handling, system specification, testing and training. This is in addition to those formally with computer science or health informatics (HI) qualifications and specific HI roles. Distinctions between competencies/knowledge experientially gained, and that achieved from formal study diminishes as careers lengthen. The challenge is to define an identity for a ‘health informatician’ that is acceptable whatever personal development route was taken.

Variations in definition from ‘computer services for health’ to specific sub-domain specialisms like ‘clinical informatics’, ‘compunetics’ [1] and ‘e-health’ arose over time, and can confound the cohesion of the domain. There are also nuances between countries, even in generic terms; from the all-inclusive medinfo2001 definition : HI is ‘..concerned with the systematic processing of data, information and knowledge in medicine and healthcare. The domain covers computational and informational aspects of processes and structures, applicable to any clinical or managerial discipline within

the health sector whether on a tele (remote) basis or not. Health informatics is delivered by operational health practitioners, academic researchers and educators, scientists and technologists in operational, commercial and academic domains' to the succinct NHS in England phraseology : '*knowledge, skills and tools that enable information to be collected, managed, used and shared to support delivery of healthcare and to promote health [and wellbeing]*' which has been adopted by the professional accreditation body, the UK Council for Health Informatics Professions [2].

There is no de facto standard term so specialist terms compound the apparent domain fragmentation. Adoption of the IMIA scientific topic map [3] and the 'Otley' definitions of core theories and concepts [4] which is currently being extended to include synonymic terms, will help to clarify what HI - by any of its names - actually embraces.

1. Discussion

1.1. Health Informatics Practitioners

The community of practice is also eclectic by virtue of many career routes into it; typically by academic qualification, experiential route after school, by qualification in another discipline then migrating into HI, or by making a conscious sea-change in career after qualifying and working in computing in a non-health domain, health management or clinical specialty. Traditional computer science routes also fragment, into specialised areas such as multi-media or AI. In view of scarcity of specialist entry-point jobs, those graduating from niche areas frequently are required to enter generic informatics in the first instance.

Trans-national movement is already observed, some based on multi-national academic courses like Erasmus [5] but as yet opportunities are limited. Standards such as [4] will stimulate much improved mobility opportunities over time.

Situations in individual countries are different; mobility depends to a large extent on local workforce positions, external attractions and personal travel aspirations. Factors explored in this paper arose from assessing the UK situation, but from international discussion [6] are known to have overall relevance.

1.2. Professionalism

Professionalism is defined by Benson [7] as operating to a code of conduct, within ethical rules, and to public benefit whilst also giving leadership in a field of learning. Professionalism also needs clarity of what constitutes 'fitness to practice', formal entry requirement(s) and a commitment to career-long continued personal development to maintain skills, knowledge and awareness of innovation in relevant technologies and the domain. Bunker [8] considers that 'self-regulation' (or 'self-policing' as Freidson [9] calls it) results in reduced individual autonomy in the medical profession. At this stage of development HI needs to take this step to contribute to maturity across the community and generate some harmonization across the range of actual roles within HI.

The UK recognized (2002) that a significant proportion of (mainly) non-clinical professionals in the NHS were involved for a substantial part of work-time in informatics-related tasks; such as information governance, knowledge management,

information analysis, training, research and operational support to care management / delivery. The UK Council for Health Informatics Professions (UKCHIP) was established to register and recognise those whose information-related activities could have a negative impact on patient safety if carried out unprofessionally. By 2006, the voluntary register became open for public scrutiny, having over 10% of the domain-active professionals on its books. In 2007, UKCHIP is developing strong employer recognition and incorporating accreditation, including of the emerging multi-professional shared Health Information Services that support care delivery, into its remit.

1.3. Meeting projected demand, addressing a potential resource gap

The eSkills Council for the information technology sector in the UK, in a recent survey [10] predicted demands for HI-competent staff escalating dramatically before 2010. Even entry level courses, addressing fundamental health informatics, such as Uclan Foundation degree [11] take two years. Realisation of benefits from the new generation of HI solutions, such as English implementations [12], is in jeopardy from the potential lack of enough skilled people to utilize the promised new functionality.

School leavers exit with extensive computing competency from study and/or the ubiquity of home computing / gaming. However clinical sensitivity of personal health data requires further training and induction in information governance, ethical handling and operating specific functional solutions. Moves such as [13] need to be put in place now, so a larger HI professional pool becomes available to avoid expanding domain-knowledgeable resource gaps mid-term.

1.4. Synergy between sectors

Health informatics is 'multi-disciplinary' as Morris states [14]. Care providers (public and private); academic institutions who teach and develop students and also carry out research; and commercial vendors of solutions and services such as call centre support, outsourcing, facilities management and remote hosting have common interests and core requirements.

Frequently, contracts are required for additional external resources to provide specialist services and manpower. These resources must have domain sensitivity and knowledge in order to ensure that informatics is provided as non-invasively and effectively as possible. All staff including those employed by external contractors must demonstrate an understanding of how the health domain works. UKCHIP[2] exemplifies recognition of such knowledge, and its principles and processes are under consideration internationally. Registration or accreditation (with strong employer recognition) can stimulate domain cohesion, create an identifiable pool of sustainable resources and drive up quality, and ultimately will bring the factions / sectors closer together, potentially under a portmanteau term 'health informatics'.

1.5. Accommodating eclectic sub-domain descriptors, such as e-Health

If HI is defined as per [2] then ‘eHealth’ can sit comfortably within it; as also proposed by Oh et al [15] even given its ‘range of meanings encompassed by the term’. Full acceptance will be indicated when it is no longer felt necessary to write ‘e-health’ but the ‘e’ is assumed as integral. Similarly ‘clinical informatics’ may characterise players as clinicians; whilst ‘bioinformatics’ shifts granularity from a patient level to a molecular or similar level and ‘compunetics’ recognises interactions between consumers and care givers and connections between computers and networking. The HI domain can be thought of as analogous to a ‘salami’ – it can be sliced in many directions and prepared in various ways but will always retain its original identity as a spicy sausage!

1.6. Risks of a collective term

The collective term proposed ‘health informatics’ may result in all projects being coloured by success or failure of generically similar activities. For example, in England, there is much tension surrounding the NHS National Programme for IT [16]; tending to subsume recognition of any other worthwhile HI from any other source; a phenomenon compounded by the media.

HI has moved from the crude original description of ‘doing computing for hospitals’ but runs a similar risk of genericising individuals, leading to public misconceptions of their specialist capabilities.

Adoption of a collective term has two main aims – to facilitate the recognition of the whole community of practice as worthwhile; respectable, recognisable and operating in a professional manner. Secondly to provide a framework to activity that will more rapidly, and on a wider basis, create an understanding of what can actually be done for and by the health domain and what are the likely outcomes downstream from informatics-related actions.

Within the UK where IT professionalism per se is also being promoted [17], distinct registration as fit to practice *in health* needs to be retained, in a similar manner to registration of every doctor under the General Medical Council [18] rather than as a gynaecologist or cardiac surgeon. Longevity of accreditation bodies like UKCHIP will be assured when the HI community achieves a critical mass.

An inclusive community who understand the sensitivity of the health sector may be disenfranchised if ‘IT people and Information Managers’ only came together; and could put patient safety at risk.

Many traditional professional registration bodies such as medicine, financial services and plumbing trades developed a large range of registrant’ services commensurate with higher fees. HI cannot offer similar until it is a recognized, mature profession with a viable (financial) basis.

1.7. Benefits of a strong brand image

As ‘domain-knowledgeable registration preferred’ is emerging in job advertisements and role specifications, those registered will have the edge over others applying for the same jobs. Utilising the generic HI term in specifications will increase the likelihood of

individuals being considered for a post which is different from their current role, especially career enhancing posts. It will also be likely to widen the quality of the pool of applicants who might consider the post.

To be a member of a distinct community is reassuring in times of tension and challenge. Peer group guidance, empathy and synergy can add considerably to professional stability (through accepted occupational standards), negotiation of career (contractual) terms and continuing personal development, especially in a workplace-based context. Communal recognition creates pride and loyalty and has been proven to generate better retention of staff [19], thereby increasing value from any organisational investment in training and development of staff who work within that organisation.

2. Actions in train

Adoption of the generic 'health informatician' will form the basis for introduction of HI in countries that currently do not have numbers or distinct identities for those involved. Countries will be able to map their requirements to those of other areas transparently, developing a common currency for expression and evaluation of skills, competences and experience. Standards of practice and job descriptions can then be shared. Even if then customised to local situations, core comparability can be maintained. Personal movement between countries will then become relatively easier.

Policy bodies will be able to profile both available resource and anticipated future requirement at an aggregate level to develop credible strategic plans based on proven skill sets and workforce descriptors. Such profiles can feed into research plans, academic course developments and vocational projections, reducing likelihood of missed targets and limited exploitation of opportunities to innovate.

Where multi-national professional groups (e.g. IMIA and its regional groups or the Council of European Professional Informatics Societies and its national learned societies[20]) operate collectively, scientific mapping of concepts and the professional specification of requirements can over time be harmonised.

Academic institutions are also playing their parts in developing HI – with further education offering clear pathways from entry to higher education; with formal entry/exit points along the pathways.

3. Conclusions

Once sharable descriptors of competence, workforce requirements and standards of practice are rolled out, it will be possible to distil a distinct brand identity for health informatics not existing presently. In a rapidly changing world we cannot passively wait for this to emerge. Cooperative working can develop, share and agree the metrics that define our domain; in parallel to initiatives to recognise ourselves as professionals.

In such a collaborative manner, a sustainable capability to deliver technology-based solutions enhancing efficiency, effectiveness and efficacy of healthcare delivery can be preserved. Continued international activity provides a cohesive impetus necessary for this focus to be maintained.

Acknowledgments

This work stems from collaboration with the Board of UKCHIP and BCS HI Forum, UCLAN colleagues and the Health Informatics community internationally.

References

- [1] Carroll, D., The International Council on Medical and Care Compunetics: Executive Overview, BCSHIF Working Paper (2004)
- [2] UK Council for Health Informatics Professions www.ukchip.org
- [3] Lorenzi, NM., IMIA's Working Groups and Special Interest Groups: Connecting the World's Medical Informatics Working Groups. IMIA Yearbook of Medical Informatics 2002 pages 71-72
- [4] Murray, P., Educational Steps www.health-informatics.org/hc2005/Education.pdf
- [5] Erasmus International masters course
www.connectingforhealth.nhs.uk/delivery/serviceimplementation/modernisation/etd/iinformatics/awards
- [6] Panel Session - S080 Global Standardisation of Certification requirements and Fostering Management Specialists for Electronic Health records Systems.
- [7] Benson, Lord., Council for Professions Supplementary to Medicine, Criteria for a Group to be Considered a Profession (1992)
- [8] Bunker, JP., Can professionalism survive in the marketplace? BMJ 1994; 308:1179-1180 (07 May)
- [9] Freidson, E., Profession of Medicine: a study of the sociology of applied knowledge, with a new afterword. University of Chicago Press (1988)
- [10] eSkills Council. IT Insights: Trends and UK Skills Implications www.eskills.com (2004)
- [11] UCLAN. Foundation Degree in Health Informatics
www.uclan.ac.uk/facs/health/lshpm/healthinformatics/courses/modules/Foundation/index.htm
- [12] NHS in England. National Programme for IT www.connectingforhealth.nhs.uk
- [13] NHS Informatics Leadership and Management pilot, working document (2007)
- [14] Morris TA, McCain KW The Structure of Medical Informatics Literature J Am Med Inform Assoc. Sept-Oct; 5(5) 448-466 (1998)
- [15] Oh, H., Rizo, C., Enkin, M., Jadad, A., What is eHealth (3): A Systematic Review of Published Definitions J Med Res 2005; 7(1): e1 <http://www.jmir.org/2005/1/e1>
- [16] UK Commentators, for example re pathology,
<http://www.e-health-insider.com/news/item.cfm?ID=2288>
- [17] British Computer Society Professionalism initiative
<http://www.bcs.org/server.php?show=conMediaFile.2191>
- [18] General Medical Council. www.gmc-uk.org
- [19] Derry, B., Haw, A., Hughes, P., ASSIST: For the profession to be formally recognised, we need a clearer idea of what constitutes healthcare informatics BJHC&IM Vol 23 No 9 (October 2006)
- [20] Council of European Professional Informatics Societies www.cepis.org

All websites checked for accuracy October 2007

Address for correspondence

Dr Jean Roberts, Health Informatics Unit, LSHPGM, Room 243, Harrington Building, University of Central Lancashire, Preston, PR1 2HE, UK JRoberts1@uclan.ac.uk

Application of the Multi-disciplinary Thematic Seminar Method in two Homecare Cases - A comparative study

Isabella SCANDURRA^{a,1}, Maria HÄGGLUND^a and Sabine KOCH^{a, b}

^a *Centre for eHealth, Uppsala University, Uppsala, Sweden*

^b *LIME, Karolinska Institute, Stockholm, Sweden*

Abstract. A significant problem with current health information technologies is that they poorly support collaborative work of healthcare professionals, sometimes leading to a fragmentation of workflow and disruption of healthcare processes.

Objective: This paper presents two homecare cases, both applying multi-disciplinary thematic seminars (MdTS) as a collaborative method for user needs elicitation and requirements specification. *Methods:* This study describes the MdTS application to elicit user needs from different perspectives to coincide with collaborative professions' work practices in two cases. *Results:* Despite different objectives, the two cases validated that MdTS emphasized the "points of intersection" in cooperative work. Different user groups with similar, yet distinct needs reached a common understanding of the entire work process, agreed upon requirements and participated in the design of prototypes supporting cooperative work. *Conclusion:* MdTS was applicable in both exploratory and normative studies aiming to elicit the specific requirements in a cooperative environment.

Keywords. Needs Assessment, Human Factors, Change Management, Organization, Process, User-Computer interface, Collaboration

Introduction

Homecare staff requires both information at the point of care and insight into other care professionals' work and documentation [1]. To develop ICT systems that actually support cooperation and coordination of homecare work, there is a need for a development process that focus on the entire care process and all professions involved [2, 3], as opposed to conventional methods working with one care profession at a time.

To meet the demands from integrated healthcare organizations for design of new ICT systems, we have developed a multi-disciplinary thematic seminar (MdTS) method. It differs from conventional requirements engineering methods as it implies a thorough investigation of the entire interdisciplinary cooperative work, *and* it supports

¹ Corresponding Author: Isabella Scandurra, Centre for eHealth, Uppsala University, University Hospital 82:1, SE-75185 Uppsala, Sweden. Isabella.Scandurra@ehealth.uu.se

rapid transformation of the agreed user requirements into technical specifications in order to develop appropriate ICT for the variety of occurring work situations.

To validate the method, MdTS was applied in two different homecare cases. This paper describes the various contexts and analyzes the differences of the application of the method.

1. Methods

The first author participated in both cases and could therefore describe the interventions and the contexts in which the interventions occurred [4] as an *intrinsic case study* (when the researcher has an interest in the case) [5] (cited in [6]). Sources relevant to the study were participant and direct observation, documentation and physical artifacts [4] e.g. the prototypes developed in the cases.

The MdTS method is based on established theories from the Human-Computer Interaction fields of Participatory Design (PD) [7, 8] and Computer Supported Cooperative Work (CSCW) [2, 9], and is thoroughly explained in [10]. In short, MdTS consists of a seminar series of twelve seminars with set themes (figure 1) and work was to be performed in multi-disciplinary working groups.

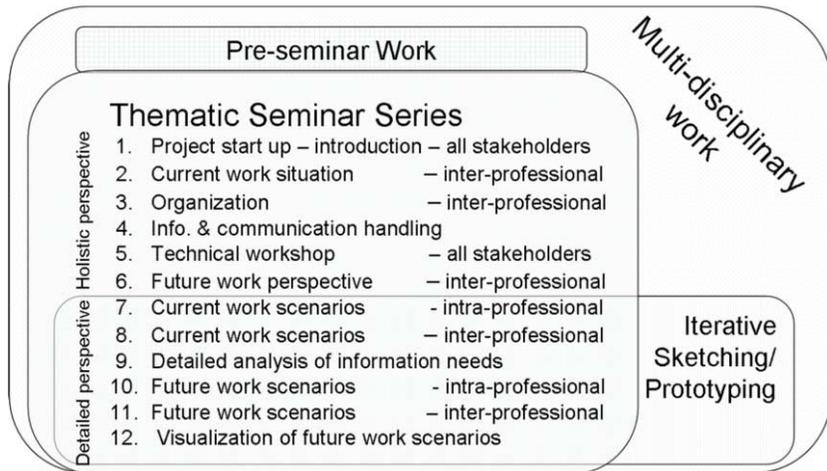


Figure 1. Thematic seminars performed in multi-disciplinary working groups.

Two *types* of multi-disciplinary seminars were held; (1) *intra-professional seminars* where focus was on one healthcare profession, and (2) *inter-professional seminars* where the cooperative aspects of work were handled. Two *perspectives* of the multi-disciplinary work were covered in the seminars; (1) a *holistic perspective* aiming to reach a common understanding among the participants and a holistic view of the entire cooperative work process, and (2) a *detailed perspective* in which results from the holistic perspective were further elaborated. The detailed perspective included examining “points of intersection” and specific details for each profession, items that were necessary for development of a health information system supporting integrated care.

The seminar series was preceded by *pre-seminar work* in form of field observations and interviews [11] carried out by health informaticians with knowledge in usability issues (i.e. HI-U specialists) [10, 12] at different workplaces to gain an understanding of the work performed and to capture users' tacit knowledge.

Analyses of the users' needs and *translations* into more technical specifications were performed by the HI-U specialists after every theme handled in the seminars. The structured analyses of the requirements were iteratively fed back to the multi-disciplinary working groups and agreed upon, as it was essential to validate the system specification during evolution. After every detailed seminar, the validated results, in the form of task analyses and use cases, were handed over to the developers.

2. Results

The MdTS method was developed in case 1, OLD@HOME, [13] and applied as an explorative study. All 12 themes were realized. In case 2, VIHO [14], aiming for normative results, the six holistic themes were applied (Table 1).

Table 1: Comparison between the cases, OLD@HOME and VIHO

	Case 1: OLD@HOME	Case 2: VIHO
<i>Total project time</i>	Fall 2002- fall 2005	Fall 2003- spring 2004
<i>Type of study</i>	Explorative	Normative
<i>Pre-seminar work</i>	17 days during 3 months, working within 3 professions	-
<i>Thematic seminar series</i>	6 months	6 months
<i>Seminars performed</i>	Holistic: No 1-6, Detailed: 7-12	Holistic No 1-6.
<i>Hours/theme</i>	Holistic perspective: 1 day/theme, 8h Detailed perspective: Each group worked for 4h/theme, c.f. iterations. Technical workshops lasted 2 days.	Holistic perspective: 1 day/theme c.f. iterations. Technical workshops lasted 2 days.
<i>Frequency of seminars and meetings between different professions in the project</i>	Seminars ~every two weeks "Homework" assignments in between, HI-U's also had an office next to HHS	Seminars ~once a month "Homework" assignments from one seminar to another.
<i>Iteration of themes</i>	- No 9; Detailed information needs - Topic: integrated care planning	No 6; Future work perspective was repeated.
<i>Field work during the seminar period</i>	6 days, 2/each profession	2 days, one day with the home care nurse, one day with HHS
<i>Participants in multi-disciplinary groups (Start-up and progress presentation meetings)</i>	Participants from stakeholders as buyers/owners, project managers, operational services, (HI-U) designers, developers and users [15]	Participants from stakeholders as buyers/owners, (HI-U) designers and users were present
<i>Participants in inter-professional working groups for holistic themes (same people during the entire seminar series)</i>	One group containing: 3 HHS (i.e. assistant nurses), 2 DN, 1 GP, 1 care planning specialist, 2 HI-U specialists. Other stakeholders received continuous progress reports.	One group containing 3 HHS, 1 home care nurse and 2 assistant nurses in elderly care, 1 HI-U specialist, 2 HCI researchers, 1 team work/organization specialist. Buyers/Owners received progress reports.
<i>Participants in Inter- & Intra professional working groups for detailed themes</i>	3 professional groups; 2 DN / 3 HHS/ 2 GPs work together (or separately) guided by 2 HI-U specialists	-
<i>Overall outcome</i>	Evaluated working prototypes, a final prototype: the Virtual Health Record.	Prerequisite documentation for requirements specification

In Case 1, **OLD@HOME**, the objective was to develop prototypes for inter-organizational ICT systems in homecare in order to provide a seamless and consistent information and communication flow between different professions caring for elderly living in private homes. Two groups of care professionals from different care providers jointly participated in the seminars. Those were general practitioners (GP) and district nurses (DN) in a *primary care* group from the County Council of Gävleborg, and a *home help service* (HHS) group i.e. 14 assistant nurses from the municipality of Hudiksvall. Three of them participated in the seminar series, and the entire group tested the prototypes during a period of five months [13]. The primary care group consisted of seven members, GP and two DN participated in the seminar series while an additional four others (two DN and two GP) participated in testing the prototypes. In the inter-professional seminars, participants from all user groups were involved.

The researchers (i.e. HI-U specialists) conducted all pre-seminar work, seminars and transformation of user needs and agreed requirements into use cases and technical specifications. Furthermore, a technical group was involved in coordinating technical infrastructure, architectural design, software development, technical implementation, and support.

The resulting system was evaluated in different ways, e.g. in a usability lab and through heuristic evaluations [16, 17]. In short, results indicated that information needed at the point of care was available to the users and presented in an understandable manner. Consequently, the MdTS method succeeded in elicitation of the correct user needs and in transferring the requirements to the system developers [1].

Case 2, **VIHO** in Kortedala, Göteborg, was a normative study, using a perspective of five years for envisioning work practices in homecare. Working ICT prototypes were not expected as result, but a proposal of future elderly care in terms of *aspects of future work* [18]. The VIHO working group consisted of five assistant nurses (three HHS and two working in an elderly care centre), and one homecare nurse. Four researchers, two HCI (human computer-interaction) specialists, a HI-U specialist and an expert in team work and organizations, provided input to the discussions and compiled the results for further reflections and analyses of the aspects of future work.

2.1. Analysis

Both projects resulted in valuable documentation on how cooperative work in homecare is currently performed and how it should preferably be performed in the future, as well as specific and detailed needs and requirements regarding information and communication handling in integrated homecare.

In case 1, extensive fieldwork was carried out, 17 pre-seminar days during 3 months containing work within three professions. Additionally, the HI-U specialists obtained an office in the HHS building and consequently spent an extensive amount of time with the staff, sharing their daily work problems at informal meetings or coffee breaks. This fieldwork was of great value for both projects. Understanding of current work situations, knowledge about problems and difficulties in cooperative work, as well as staffs' thoughts about their work created a common ground and facilitated for communication between HI-U specialists, researchers and healthcare participants in both cases.

In both cases, the MdTS was conducted during a period of 6 months. Case 1, OLD@HOME, used the MdTS method to acquire and analyze user requirements in cooperative work with the objective to develop a virtual health record (VHR). As the results were iteratively refined in the second part of the seminar series, a working prototype of the VHR for integrated homecare was developed, allowing for information access and documentation at the point of care. [13]

In case 2, VIHO, the seminar work was considered finished when a prerequisite for requirements specifications for future work in homecare was accomplished. Case 2 iterated seminar 6, *Future work perspective* to finalize the work accordingly. Analyses were performed out of results from the holistic seminars and no analyses focused on details in order to *develop* a system, as in OLD@HOME. Accordingly, the total work load was lower in VIHO; staff participated in six holistic seminars and spent one day/theme. Case 1 needed more detailed descriptions about the information needs (no 9) and how the care planning process was performed. As required, case 1 performed much more detailed analyses. User requirements elicited in the holistic seminars were transformed into VHR prototypes and the conformity was verified in user tests, performed during the detailed phase of the seminar series. Consequently, additional time (4h/detailed seminar) was demanded.

When considering details in daily work activities studied in the two cases, the tasks contained few differences and on a general level the user needs specifications were in accordance. As a second case, VIHO validated the method developed in OLD@HOME and verified the results from the MdTS by briefly testing the prototypes developed in case 1. During development and evaluation of the prototypes, several flaws were found and adjusted, thoroughly described in [1].

3. Conclusions

MdTS is specifically adapted for cooperative work, to support gathering of requirements from different perspectives to coincide with different staff's work practices. However, we consider the MdTS method to be generally applicable. It can e.g. be used for single profession analysis if the inter-care professional seminars number 8 and 10 are removed, and it can be applied to other domains than healthcare.

To develop a system that supports cooperation and sharing of patient information between different care professions, both the holistic *and* the detailed part of the multi-disciplinary thematic seminars are needed. For other purposes, e.g. activity and information needs analyses or normative studies, application of the holistic part is sufficient. Application of the method resulted in a holistic view of the entire work process achieved by the various participants and specifications of goals for future cooperative work and ideas of adequate ICT support. The *detailed* part specifically took into consideration the *inter-care professional aspects of work* and focused on details needed to develop usable HIS for cooperative work. The detailed phase was consequently regarded as mandatory as the objective was to develop an ICT tool, to ensure that it actually supported integrated care practices.

Despite different objectives, the two cases demonstrated that the method emphasizes the "points of intersection" and benefits both healthcare participants from different professions *and* system developers in the delicate phase of eliciting user needs and creating system requirements as well as ICT tools for integrated care.

Acknowledgements

The project “OLD@HOME - Technical Support for Mobile CloseCare” (P23037-1 A) was supported by VINNOVA - Swedish Agency for Innovation Systems, Trygghetsfonden and by its clinical and industrial partners. The project “VIHO - Efficient computer support in care for the elderly” was sponsored by VINNOVA (2001-05296).

References

- [1] Scandurra I. Building usability into health informatics. Development and evaluation of information systems for shared homecare. Uppsala: Uppsala University; 2007.
- [2] Pratt W, Reddy MC, McDonald DW, Tarczy-Hornoch P, Gennari JH. Incorporating ideas from computer-supported cooperative work. *Journal of Biomedical Informatics*. 2004;37:128-37.
- [3] Hartswood M, Procter R, Rouncefield M, Slack R. Making a Case in Medical Work: Implications for the Electronic Medical Record. *Computer Supported Cooperative Work (CSCW)*. 2003;12(3):241-66.
- [4] Yin RK. *Case study Research: Design and methods*. Thousand Oaks, CA: Sage Publications; 1994.
- [5] Stake RE. *The art of case study research*. Newbury Park, CA: Sage Publications; 1995.
- [6] Tellis W. Application of a case study methodology. *The Qualitative Report* [On-line serial]. 1997 September 1997;3(3):<http://www.nova.edu/ssss/QR/QR3-/tellis2.html>.
- [7] Clemensen J, Larsen S, Kyng M, Kirkevold M. Participatory Design in health sciences: using cooperative experimental methods in developing health services and computer technology. *Qual Health Res*. 2007;17(1):122-30.
- [8] Bødker S, Ehn P, Sjögren D, Sundblad Y. Cooperative design perspectives on 20 years with "the Scandinavian IT Design Model". In: Gullikssen J, editor. *NordiCHI 2000, Design vs Design*; 2000; Scand.
- [9] Hardstone G, Hartswood M, Procter R, Slack R, Voss A, Rees G. Supporting informality: team working and integrated care records. *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work 2004*; Chicago, Illinois, USA, November 06 - 10, 2004: ACM Press; 2004. p. 142-51.
- [10] Scandurra I, Hägglund M, Koch S. From user needs to system specifications: Multi-disciplinary thematic seminars as a collaborative design method for development of Health Information Systems. *Journal of Biomedical Informatics*. 2007;Accepted.
- [11] Beyer H, Holtzblatt K. *Apprenticing with the Customer: A Collaborative Approach to Requirements Definition*. Communications of the ACM. 1995 May.
- [12] Samaras GM, Horst RL. A systems engineering perspective on the human-centered design of health information systems. *Journal of Biomedical Informatics*. 2005;38(1):61-74.
- [13] Koch S, Hägglund M, Scandurra I, Moström D. *Old@Home - Technical support for mobile closecare*, Final report; 2005. Report No.: VR 2005:14.
- [14] Hardenborg N. *Designing Work and IT systems - a Participatory Process that Supports Usability and Sustainability*. Uppsala: Acta Universitatis Upsaliensis; 2007.
- [15] IEEE Architecture Working Group I. *Recommended Practice for Architectural Description of Software-Intensive Systems*: IEEE Std 1471-2000, IEEE; 2000.
- [16] Scandurra I, Hägglund M, Moström D, Koch S. Heuristic Evaluation Extended by User Analysis; A fast and efficient method to identify potential usability problems in Health Information Systems. *Journal on Information Technology in Healthcare*. 2006;4(5):317-25.
- [17] Scandurra I, Hägglund M, Koch S, Lind M. *Usability Laboratory Test of a Novel Mobile Healthcare Application with Experienced Homecare Staff*. to be submitted. 2008.
- [18] Johansson N, Scandurra I, Eriksson L, B S. *Aspekter på det framtida arbetet inom äldreomsorgen*. Uppsala: Uppsala university, Dept of Information Technology; 2005.

8. National eHealth Roadmaps, Cross-Border Applications and Organisational Strategies

This page intentionally left blank

Modelling access to renal transplantation waiting list in a French healthcare network using a Bayesian method

Sahar BAYAT^a, Marc CUGGIA^a, Michel KESSLER^b, Serge BRIANÇON^c,

Pierre LE BEUX^a, Luc FRIMAT^c

^a*EA 3888, Université Rennes 1, IFR 140, Rennes, France*

^b*Service de néphrologie, Hôpitaux du Brabois, Nancy, France*

^c*EA 4003, Nancy Université, France*

Abstract: Evaluation of adult candidates for kidney transplantation diverges from one centre to another. Our purpose was to assess the suitability of Bayesian method for describing the factors associated to registration on the waiting list in a French healthcare network. We have found no published paper using Bayesian method in this domain. Eight hundred and nine patients starting renal replacement therapy were included in the analysis. The data were extracted from the information system of the healthcare network. We performed conventional statistical analysis and data mining analysis using mainly Bayesian networks. The Bayesian model showed that the probability of registration on the waiting list is associated to age, cardiovascular disease, diabetes, serum albumin level, respiratory disease, physical impairment, follow-up in the department performing transplantation and past history of malignancy. These results are similar to conventional statistical method. The comparison between conventional analysis and data mining analysis showed us the contribution of the data mining method for sorting variables and having a global view of the variables' associations. Moreover these approaches constitute an essential step toward a decisional information system for healthcare networks.

Keywords: Data analysis-extraction tools; Decision support; Organization; Bayesian network; Healthcare network; renal transplant waiting list

Introduction

Incidence and prevalence of end-stage renal disease (ESRD) requiring renal replacement therapy (RRT), i.e. hemodialysis, peritoneal dialysis or kidney transplantation, are still increasing [1]. Kidney transplantation is associated with longer survival and lower long-term cost [2]. But, given the graft shortage, transplantation with a cadaveric donor kidney before start of RRT is not commonly achievable. On the other hand, all patients are not equally suited for transplantation and there is a body of evidence showing that selection criteria of the potential transplant recipient diverges from one centre to another [3]. Ideally, placement on the waiting list should be based solely on medical factors, in accordance with medical guidelines. However, previous

studies showed that blacks, Hispanics, women and elderly were less likely to receive a renal transplant in some countries. Moreover, distance from transplantation department, and private ownership of dialysis facilities have been associated with poor access to kidney transplant waiting lists [4-6]. The network of care NEPHROLOR was set up in the French administrative region Lorraine to improve quality of care of all patients with chronic kidney disease (CKD). NEPHROLOR has set up a Regional Health Information System in order to collect patients' data at first RRT and during the follow up. These data are potentially useful for optimizing organization in NEPHROLOR. Indeed, Access to renal transplant waiting list is one key point to achieve optimal care. The purpose of the present study was to evaluate the feasibility and suitability of Bayesian networks for detecting factors associated with registration on the waiting list. These methods have been already used to predict survival of liver[7] or lung [8] transplant patients but we have found no published study using Bayesian approach to evaluate access to renal transplantation.

1. Material and method

1.1. Organisation of care for CKD in Lorraine region

The NEPHROLOR network of care was set up, in Lorraine, in 2002. It combines the public and private for-profit facilities, operating dialysis units in Lorraine. In Lorraine as in most French regions, renal transplantation is performed at the University hospital nephrology department which is the only transplant centre of the region.

1.2. Study population

Since June 1997, all adult patients living in Lorraine and starting RRT (dialysis or pre-emptive transplantation) in a NEPHROLOR facility were progressively registered in a regional database [1, 4]. For the present study, we included the incident patients between July 1, 1997 and June 30, 2003. The patients having missing data were not included. In order to identify all ESRD patients living in Lorraine and placed on the waiting list, the list of patients registered on the French national waiting list between January 1, 1996 and December 31, 2004 was extracted from the CRISTAL[9] database and the list of transplanted patients from the database of the transplantation department of Lorraine.

1.3. Data collection

For each inclusion of a new patient in the NEPHROLOR database, a standardized form is prospectively filled out at the initiation of RRT [1, 4]. Three categories of variables possibly related to registration on the transplant waiting list were studied.

The first included social and demographic data: age, sex and residence at first RRT. The distance between the patient's residence and the department performing transplantation was calculated in kilometers. In France, legal regulations prohibit considering ethnic differences in the French ESRD registry. However, we know that, in Lorraine, almost all patients were Caucasians, with a significant proportion of natives from Italy, and North Africa. In the French ESRD registry, descriptions about income

or education were not available [1]. The second category included clinical, anthropometric and biological data at first RRT: existence of diabetes, cardiovascular disease (coronary artery disease, peripheral vascular disease, congestive heart failure and cerebrovascular disease), respiratory disease, hepatic disease, psychiatric disorder (severe depression or other psychiatric disorder) and past history of malignancy. Patients who were confined to a wheelchair or were bedridden were considered to have physical impairment of ambulation. Body Mass Index was categorized in <20 kg/m², 20-24.99, ≥ 25 , hemoglobin in <11 g/dl, ≥ 11 , and serum albumin in <3 g/dl, 3-3.49, ≥ 3.5 . The third category included data related to medical follow up in the NEPHROLOR network: ownership of nephrology facility where the first RRT was performed: public, or private. We also took into account the effect of a medical follow-up in the department performing transplantation versus 12 other facilities without transplantation.

1.4. Statistical analysis

As the oldest patient registered on the waiting list was 78 years old, we decided to exclude patients older than 80 years on the first day of RRT from statistical analysis. A conventional statistical analysis was done using descriptive method and statistical tests. Impact of baseline characteristics on registration on the waiting list was univariately analyzed using the χ^2 test. Subsequently, all variables univariately significantly associated with registration with p of 0.10 or less were presented stepwise to a multiple logistic regression model to assess their independent value for registration. Within each step, significant risk factors were selected with a forward strategy using the likelihood ratio statistic, with p of 0.05 on the criterion level of selection. These analyses were performed with SAS software (version 9.1; SAS Institute Inc, Cary, NC). We performed also a Bayesian network [10] analysis with BAYESIALAB Software (version 3.3; Bayesia SA). A Bayesian network describes a system of interest by specifying relationships of conditional dependence between its variables. These relationships are represented by a graph, in which, the nodes represent the variables and the arcs represent relationship between variables. The thickness of the arc between 2 variables is proportional to the contribution of their relationship in characterization of the network. The brightness of square represented in a node is proportional to the contribution of this node in the characterization of the target. The Bayes' rules can be described with the following formula:

$$P(R = r|e) = \frac{P(e|R = r)P(R = r)}{P(e)}$$

Where $P(R = r | e)$ denotes the probability that random variable R has value r given evidence e . In these models, a simple representation of a complex problem is possible. The analysis were performed in three steps: learning step (SopLEQ method)[11-13], analysis of associations, inference in order to characterize registration on the waiting list (target node). We used the parent-sons model to characterize the target node. These models allow us to deduce which variables are directly related with the target node.

2. Results

2.1. Patients' characteristics

Eight hundred and nine patients were included in the study. Their mean age was 62.1 ± 14.2 years; 482 (59.6%) were male, 279 (34.5%) had diabetes, 358 (44.2%) cardiovascular disease and 90 (11.1%) respiratory disease. One hundred and fourteen (14.1%) had a past history of malignancy, 158 (19.5%) physical impairment and 48 (5.9%) psychiatric disease. Two hundred and twelve (26.2%) of them were registered on the transplant waiting list.

2.2. Logistic regression results

Table 1 presents factors associated with registration on the waiting list using a logistic regression model. The younger the patient, the more frequent was the registration on the waiting list. For example, compared with a patient older than 65 years, a patient younger than 44 years was 39 times more likely to be placed on the waiting list. Among co-morbidities, psychiatric disorders, cardiovascular disease, diabetes, past history of malignancy, and respiratory disease were independent factors associated with non registration. Patients with serum albumin greater than 3.5 g/dl were 2.7 times more likely to be placed on the waiting list, compared to those having serum albumin less than 3 g/dl. Patients followed in the nephrology department performing transplantation were 85% more likely to be registered.

Table 1. Factors associated with registration on the waiting list (multivariate logistic regression analysis, n= 809 patients). For the co morbidities, presence of pathology represents the reference modality.

Characteristics	OR	95% CI
Age categories		
≥ 65 years	1	
55 – 64	11.50	6.10 – 21.68
45 – 54	33.00	16.62 – 65.53
18 – 44	38.99	18.57 – 81.89
Psychiatric disorder	7.37	2.23 – 24.37
Cardiovascular disease	3.26	1.89 – 5.61
Serum Albumin		
< 3 g/dl	1	
3 – 3.49	1.54	0.76 – 3.11
≥ 3.5	2.74	1.46 – 5.16
Diabetes	2.97	1.67 – 5.28
Past history of malignancy	2.73	1.29 – 5.79
Respiratory disease	3.50	1.16 – 10.56
Medical follow-up in the department performing transplantation		
No	1	
Yes	1.85	1.15 – 2.98

2.3. Bayesian model results

For the learning step, we used the SopLEQ method to build the network (Fig. 2) The graph of the variables showed that the probability of being registered on the waiting list (target node) is directly associated with 6 variables: age, diabetes, cardiovascular disease, past history of malignancy, serum albumin level and follow up in the department performing transplantation. The Parent-Son relationship (Fig. 3) demonstrated that one node appears to be highly related to the target node. This node is the ‘Age’ node. Diabetes, cardiovascular and respiratory diseases, physical impairment, serum albumin level, past history of malignancy and medical follow-up in the department performing transplantation were also related to registration on the waiting list. Relative contribution of the variables to characterization of the target node were as following: Age: 1.0000; Cardiovascular disease: 0.3387; Diabetes: 0.2016; Albumin: 0.0837; Respiratory disease: 0.0762; Physical impairment: 0.0504; Follow up in transplantation center: 0.0423; Past history of malignancy: 0.0395.

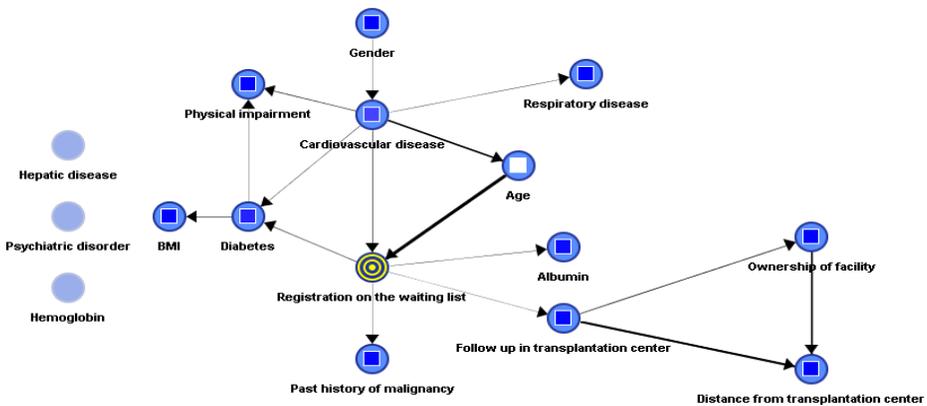


Figure 2 - Bayesian Network (SopLEQ learning method)

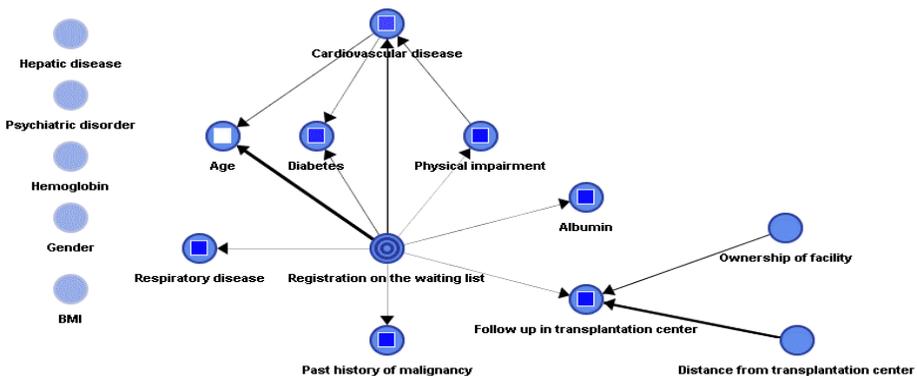


Figure 3 – Parent-Son relationship

3. Discussion

The results of the conventional analyses and the Bayesian methods were similar. It appears that in Lorraine, access to transplant waiting list is primarily associated with medical determinants. Differences due to sex, ownership of dialysis facility, and distance to transplant centre that have been identified in other studies [5, 6] were not found in Lorraine region. Conventional statistical analyses enable the clinician's hypothesis to be tested and this can produce results with statistical and clinical significance notions. Conversely, data mining methods allow a complex universe to be untangled. While, data mining methods represent a global view of relationships between variables instantaneously, it is necessary to spend much more time to obtain the same conclusion with the statistical analysis. Moreover, the relationships between variables are provided without prior hypothesis. The sample of data 'speaks for itself' and illustrates alone which variables are associated with the others. The inference power offered by Bayesian network is well accepted by physicians and comprehensive. Furthermore, it allows the experts to imagine other relationships or to test impact of events for a diagnosis or a disease.

4. Conclusion

Bayesian networks allowed us to describe the determinants of access to renal transplant waiting list. Furthermore, compared to conventional statistical method, it provides a global view of variables associations. Data mining constitutes an essential step toward a decisional information system for healthcare networks. Indeed, the result of our study can be used for optimizing renal transplant registration process in NEPHROLOR.

References

1. Couchoud, C., et al., The renal epidemiology and information network (REIN): a new registry for end-stage renal disease in France. *Nephrol Dial Transplant*, 2006. 21: 411-8.
2. Wolfe, R.A., et al., Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. *N Engl J Med*, 1999. 341:1725-30.
3. Jeffrey, R.F., et al., Comparison of transplant listing strategy in two renal dialysis centers within a regional transplant alliance. *Clin Nephrol*, 2005. 64:438-43.
4. Kessler, M., et al., Impact of nephrology referral on early and midterm outcomes in ESRD: EPidemiologie de l'Insuffisance RENale chronique terminale en Lorraine (EPIREL): results of a 2-year, prospective, community-based study. *Am J Kidney Dis*, 2003. 42:474-85.
5. Oniscu, G.C., et al., Equity of access to renal transplant waiting list and renal transplantation in Scotland: cohort study. *BMJ*, 2003. 327:1261.
6. Garg, P.P., et al., Effect of the ownership of dialysis facilities on patients' survival and referral for transplantation. *N Engl J Med*, 1999. 341:1653-60.
7. Hoot, N. and D. Aronsky, Using Bayesian networks to predict survival of liver transplant patients. *AMIA Annu Symp Proc*, 2005:345-9.
8. Smits, J.M., et al., Predictors of lung transplant survival in eurotransplant. *Am J Transplant*, 2003. 3(11):1400-6.
9. Strang, W.N., et al., The French organ transplant data system. *Stud Health Technol Inform*, 2005. 116:77-82.
10. Goldszmidt, M. and N. Friedman, *Learning Bayesian Network from Data*. SRI International, 1998.
11. Jouffé, L., Nouvelle classe de méthodes d'apprentissage de réseaux bayésiens. *Journées francophones d'Extraction et de Gestion des Connaissances (EGC)*, 2002.
12. Jouffé, L. and P. Munteanu, New Search Strategies for Learning Bayesian Networks. *Proceedings of Tenth International Symposium on Applied Stochastic Models, Data Analysis*, 2001.
13. Munteanu, P. and M. Bendou, The EQ Framework for Learning Equivalence Classes of Bayesian Networks. *First IEEE International Conference on Data Mining (IEEE ICDM)*, 2001.

Evaluation of Robustness of a User Requirements Specification Approach in a Purchase Context, a LIS Case Study

Jytte BRENDER^a and Peter McNAIR^b

^a *Department of Health Science and Technology, Aalborg University, and V-CHI, Aalborg, Fredrik Bajersvej 7D, 9220 Aalborg E, Denmark; Jytte.brender@v-chi.dk*

^b *Health Informatics Unit, Koncern IT, Capital Region of Denmark, Copenhagen NV, Denmark.*

Abstract: *Objectives:* To verify the usefulness in terms of robustness of a specific User Requirements Specification Document (URD) approach in a contractual context. *Methods:* The URD was specified in the users' terminology and in a hierarchical structure (strategic, tactical and operational requirements). The organisational setting concerns the purchase of a LIS for 6 laboratories, anticipated being taken into operation May 2008. *Results:* An immediate description of the observations is that: i) amazingly few contract amendments or other changes have been necessary, ii) the budget is kept except for the labour resources within the user organisations, caused by the vendor's delayed final delivery, and iii) the functionality at present is judged to fulfil the functional needs of the user organisations. *Conclusion:* The recommendations were helpful in the users' quality assurance of the purchase and implementation process: The contractual foundation as well as the budget were indeed very stable despite several years of delays, presumably because it is based on a purpose-oriented and hence fairly timeless URD.

Keywords: Evaluation; Laboratory Information System; User Requirements; Needs assessment; Quality Management; Purchase; Strategic Plans.

1. Introduction

The present paper reports from a long-term follow-up evaluation study on conclusions and recommendations from an in-depth case study. The paper outlines the essence of the approach and the outcome of applying the recommendations.

The European Union is regulated by law to enhance competitiveness, implying that public institutions must have an open tender process for tenders over a certain size. Almost all dedicated IT-based solutions for the healthcare sector exceed the threshold size, thus emphasising the role of a User Requirements Specification Document (URD). This again enforces concerns related to preparation of a detailed URD and the completion of a tender, with built-in assumptions and inherent problems in formulating a URD, like i) the user's difficulties in expressing their require-

ments correctly, completely and in accordance with the technological possibilities, and ii) that things often work differently than users specify them, to some degree disabling certain user requirements approaches. See the discussion in [1], p. 21.

In the mid to late 1970'ies a major and 1st generation system development was accomplished, implementing a Laboratory Information System (LIS). The system was tailor-made based on a formal system development approach with extensive user-participation and involvement of consultants on the user site. Despite the contract and the warranty both being fulfilled, it was clear that the system had problems in operation. A thorough case study explored the causal relations behind the operational problems, through investigation of the daily operation, the historical documents of the development project, and the organisational and contractual relations that led to the end result. The conclusion was that a major share of the main problems identified was closely related to the approach and role of the URD, [2]. The subsequent recommendations were among others: 1) the URD should be prepared in the users' terminology and on their conditions; 2) The URD should be purpose-oriented (rather than solution-oriented) and structured according to strategic, tactical and operational aspects. We have previously shown that preparing a URD in this way [3] is a huge advantage in the preparatory URD phase and in the selection of the vendor; see [4] and [5]. The objective of the present study therefore is to verify the usefulness of the recommended approach as regards how the URD approach works in a contractual context. In the present study this is concerned with whether the URD was sustainable (robust) – that is, whether or not the URD could cope with the given conditions for the case at hand.

This follow-up study reported is concerned with the implementation of the 3rd generation LIS at the same premises as the first LIS, but now meant to serve 6 federated laboratories for 6 hospitals in a corporation of hospitals, and with the chief physician of the original case study laboratory as chairman of the steering committee: An ideal case for assessing the validity of the recommendations in [2].

2. Study Context

The first author served throughout the purchase project in an action-research case role as consultant, however subordinate to and somewhat marginalised by the project management, and hence the role compares somewhat to the second approach to action research outlined in [6]: the researcher operates through informed action and reflection, but with no control of the process. The second author served from the start until mid summer 2007 as chairman of the Steering Committee of laboratory directors, etc., in close collaboration with the project management.

2.1. Organisational Setting

The health organisation where the system is being implemented comprises six independent hospitals and their biochemistry laboratories in the Capital Region of Denmark (in short: RegH). Two of the laboratories are accredited according to ISO 15189, one according to ISO 9001, and five of the hospitals are accredited according to the Joint Commission on Accreditation of Healthcare Organisations. Especially the ISO 15189 enforces special demands as regards the quality control of

the LIS system itself, its ability to operate with the laboratory practice in general, as well as on the relation between the laboratories and their vendor or subcontractors.

2.2. *The LIS*

The LIS project has the purpose of replacing the existing five LIS for the six hospitals in RegH, essentially with the same functionality as the old LISs: 1) CPOE functionality to support online requesting of laboratory service orders as well as their reporting within the clinics and ambulatories. 2) Within the laboratories, the LIS supports and enables the entire laboratory production from the ordering to phlebotomy, over the analytical production to quality management of the analytical production, and to securing appropriate reporting. General practitioners and collaborating commercial laboratories communicate requests and reports through EDIFACT messages directly with the LIS.

The call for tender was international and aimed at purchasing an existing product. Only two of the six bids were from Danish or Scandinavian vendors. Site visits and a dedicated investigation showed that the functionality of a LIS is fairly common in the Western culture. The LIS purchased is an existing, commercially available LIS product that has dominated the Danish market for some 20 years. However, the version offered comprises a re-implementation of that LIS on an up-to-date technical platform. The delivery of the LIS is seriously delayed because of delays in the technical development. Today the implementation process is at the end of an exhaustive long-term verification of the functionality's compliance with the contract and at the same time validation against today's practice in the 6 different laboratories. At present there are no signs predicting a failure as regards the daily operation based on the functionality known today; the main risks being technical aspects, like performance plus a few missing pieces of the functionality. In the moment of writing a final acceptance trial is planned for late February 2008 with anticipated operation of the hospitals anticipated beginning Autumn 2008.

The LIS was taken into operation mid 2006 and late 2006, respectively, at two smaller customer sites in Denmark each also with a set of hospitals and laboratories.

2.3. *The historical progression*

The purchase process started in the Hospital Corporation of the Copenhagen, H:S, but as per 1st of January 2007 H:S was merged with three counties to form RegH. One of these three counties had similarly accomplished a call for Tender shortly before the fusion, based on an almost identical User Requirements Specification, so the two contracts were merged. The third county has an old, American LIS serving three hospitals, while the fourth county has still another LIS. Organisationally, the management of the LIS project was transferred from H:S to RegH's IT Directorate.

3. **Methods and materials**

The User Requirements was prepared as recommended by the initial case study [2], for details see [3]: it was specified in the user terminology while adopting a hierarchical structure (strategic, tactical and operational requirements). This description addresses the organisation, its mission and conditions, and the functionality that the

LIS must serve at increasing details. More detail and several examples are provided in [3]. The end result was a total of 387 individual requirements.

Elicitation of the URD was carried out – roughly – in 2000; and was synthesised, extended and adapted from a handful of existing LIS URD's, of which one was from Norway (the preferred one) and one from Ireland, while the rest were from Danish colleagues. The tender process, selection of a vendor and negotiation of a contract were accomplished in 2001. The URD was discussed carefully and detailed with the vendor, item per item, resulting in addition of annotations where relevant to make further specifications or clarification from either side prior to signature. Anticipated daily operation at the first hospital was February 2003.

Evaluation of how well the URD approach served as a contractual document is measured by means of deviations from the original contractual agreement. There are two types of deviations: contractual amendments and less formal agreements. These deviations were analysed for implication and causal origin. A synthesis of the quantity and nature of these deviations constitutes the basis for judging how well the URD approach served its purpose in a contractual context.

4. Results

The total number of formal contractual amendments is 31 plus one in process to handle the latest delay. The number of less formal agreements with implications for the functionality is 20. Of the 31 formal amendments, none addresses changes or additions of the user / application-oriented functionality, except for a couple of options that were decided to become included. Most of the amendments address timing issues, and/or hardware and basic software changes to cope with performance problems, or follow-up on technological advancements (like screen resolution and Internet Explorer version).

The 20 less formal changes were a mixture of no longer relevant requirements, postponed details in the functionality, evolution of the national legislation respectively the organisational structure, and accepted constraints on narrow and specific details. Two changes were deliberately and explicitly chosen / signed away.

Moreover,

- the original budget is kept, except for labour resources within the user organisations. The anticipated cost containment inherent in the new LIS functionality was timely harvested as the vendor took over maintenance and operation of the old LISs on the originally agreed delivery date
- the user consensus as regards the functionality at present is that the functionality is anticipated to fulfil the needs of the user organisations

The LIS project is in the middle of negotiating anew with the vendor specifically with respect to timing and remaining issues like performance.

5. Discussion

A learning cycle is at its end: recommendations derived from lessons learned at an in-depth case study served as the hypothesis for verification in a second case study, in a stepwise fashion. The first step was successful as described in the Introduction. The present study deals with step two: did the URD fulfil the users need in practice

in a contractual context? It is however, a pilot study, as the final conclusion has to await the outcome of an analysis of the daily practice during operation.

The fact that the URD was purpose-oriented rather than specific in its requirements made it timeless and impervious to the change demands that usually follow some 7-8 years of evolution, technologically as well as in the application domain, including the change of two laboratories accreditation from ISO 17025 to the newly established ISO 15189. Further, our interpretation from the observations is that the nature of the URD conceivably made it too hard for the vendor to handle the customer's rights according to the contract. The URD has been used with minor modifications by three other Danish counties out of six with recent LIS tenders.

The vendor's knowledge of the domain and the specific market no doubt has influenced the extent of required changes as regards the functionality, but our own investigations have shown that the differences between laboratory practices in Europe are actually pretty small. So, this cannot explain the findings alone.

The budget was stable, and the type and relatively low number of contractual changes, indicating that the URD approach was strong – that is, the users have had no need to negotiate continuously at the level of the solution during the realisation of the project, but could support the vendor with the fine details of state-of-the-art practice that is normally below the details of an enterprise model exchanged. The reason for this stronghold may be dual: 1) the characteristics of the URD, and/or 2) the high computer literacy of the users involved, of course realizing that the two are closely related. Thus a weakness of the approach may be its generalisability: It is judged generalisable for other types of systems and other application domains with a similar type of user organisation; the weakness being that the users in this study are highly computer literate with 3rd respectively 4th generations of LIS in the organisation, so they are experienced with the functionality needed. It is therefore a question to be verified by a future project whether the URD approach is as successful for de novo IT developments and for computer illiterate users.

The strength of the study was the convincingly small number of contract amendments and the stable budget despite the prolonged implementation time. Given that the study is of a pure observational kind, with data collected over a seven-year period – interpreting what is – there is no unexpected events, little chance of confounding factors or other types of biases. The type of action-case role of the consultant literally being without control of the legal process in this study context was an advantage, because the historical process was allowed to unfold itself beyond the consultant's control. The URD/contract approach obviously is strong, but is it too strong – putting the vendor out of control of his developments, because it enables continued and elaborated interpretations, where user adaptation might work as well? These are points that need further exploration as well.

Up to date approaches used for elicitation of user requirements are enterprise modelling, use cases and scenarios, see overviews in [7] respectively [8]. The URD in the present study includes a traditional, yet high-level enterprise model, however, not of either of the two types and never explicitly dealt with within the LIS project. The difference is that the two mentioned approaches attempt to describe the (unknown) future, while the present approach describes the actual enterprise and the needs for support in this respect. So, novel is that the enterprise model serves as a supporting means rather than as the focus within the contractual relation. This is why it may not be generalisable for de novo IT developments, as discussed above. Similarly, it may be feasible only when the vendor has extensive domain insight.

Ideally, the present conclusion should be verified and further elaborated during operation of the LIS by a follow-up study of events and deviations from desired operation followed by a causal analysis of these events and deviations, the same way as the original study in [2]. Preferably this should take place at the first major hospital taken in operation, for instance, during the early (e.g. 6 months) operation.

6. Conclusion

The recommendations were helpful in the users' quality assurance of the purchase and implementation process: The contractual foundation as well as the budget were stable even over a contractual period of now almost 6 years with many delays, presumably due to the purpose-oriented approach and hence fairly timeless URD.

Authors' contributions

The first author was the main driving force as regards the study idea and its accomplishment. The second author's contribution is concerned with interpretation and judgment of the implications of the contractual amendments in a larger perspective. Both authors have approved this final version.

Competing Interests

There are no conflicting interests, financially or otherwise. There is full freedom to publish, provided that patient security issues and business secrets are kept confidential. This was achieved without consequence for the conclusion above.

References

- [1] Brender J. Handbook of Evaluation Methods for Health Informatics. New York: Academic Press, 2006. 361pp.
- [2] Brender J. Quality Assurance and Validation of Large Information Systems - as Viewed from the User Perspective, Master Thesis (Computer Science), Copenhagen University, report no 89-1-22, Copenhagen, 1989 (in Danish).
- [3] Brender J, McNair P. User requirements specifications: a hierarchical structure covering strategic, tactical and operational requirements. *Int. J. of Med. Informatics* 2001;64(2-3):83-98.
- [4] Brender J, Schou-Christensen J, McNair P. A case study on Constructive Assessment of bids to a Call for Tender. In: Surján G, Engelbrecht R, McNair P, eds. *Health Data in the Information Society, Proceedings of the MIE2002*. Amsterdam:IOS Press. *Studies in Health Technology and Informatics* 2002;90:527-532.
- [5] Brender J, McNair P. Tools for Constructive Assessment of bids to a Call for Tender - some experiences. In: Surján G, Engelbrecht R, McNair P, eds. *Health Data in the Information Society, Proceedings of the MIE2002*. Amsterdam:IOS Press. *Studies in Health Technology and Informatics* 2002;90:533-538.
- [6] McKay J, Marshall P. Action Research. A Guide to Process and Procedure. In: Remenyi D, ed. *Proceedings of the European Conference on Research Methodology for Business and Management Studies*; 2002 Apr; Reading, UK. Reading: MCIL. 2002:219-227.
- [7] Shen H, Wall B, Zaremba M, Chen Y, Browne J. Integration of business modeling methods for enterprise information system gathering and user requirements gathering. *Comput Industry* 2004;54(3):307-323.
- [8] Alexander IF, Maiden N. *Scenarios, Stories, Use Cases, Through the Systems Development Life-Cycle*. Chichester: John Wiley & Sons Ltd.; 2004.

A Vision for the Use of ICT by Norwegian Healthcare in 2012

Edgar GLÜCK, Roald BERGSTRØM, Ole-Fredrik MELLEBY and Jessica GABIN
Norwegian Centre for Informatics in Health and Social Care (KITH), Norway

Abstract. This paper presents how ICT is expected to be used by Norwegian Healthcare by the end of year 2012. The main forces influencing the ICT development are identified and the resulting set of new features is presented.

Keywords. Vision, ICT, healthcare

Introduction

As a partner with the central health authorities and as a consulting agency for the national IT healthcare sector, the Norwegian Centre for Informatics in Health and Social Care (KITH) needs to be up-to-date on the current development concerning the use of information and communication technology (ICT) in healthcare. In order to provide the best advice, we need to distinguish the forces that influence this development, establish how the various new features are interrelated and interdependent and identify any critical activities as early as possible.

For this purpose, KITH has performed a study in 2007 on how Norway will advance and integrate ICT in healthcare by the end of 2012. The expected new main features are identified and presented in this article.

1. Background

A vision is not a plan, but a visualisation of what is expected to happen. Like a plan, the final result may not be exactly as expected or occur at the scheduled time. In our case, we were to predict the status of ICT in healthcare by the year 2012.

Five years ahead is not any distant future. We would expect most of the development to be influenced by features that exist today. On the other hand, we know that revolutionary features may appear out of the sky and possibly have significant influence beyond that which we predicted.

Today, ICT is used in all parts of the healthcare environment from management of personnel and equipment to sophisticated technical instruments. For the purpose of this study, we have focused on the use of ICT in treatment processes and work flow of healthcare services more than on advances within technical equipment and management. This division is, however, not always obvious. We have also focused on the functional solutions that will be available and not on the specific technical and

organisational details that have to be in place. While ICT is important, we realize that organisational and other initiatives often are needed to achieve the full effects.

2. Method

In this study, in order to identify the upcoming new features, we have tried to identify the forces that are behind the development of ICT for use within healthcare.

2.1. Medical, Social and Political Development

New medical knowledge allows for treatment of increasingly more diseases, but such treatment typically requires the joint effort of several specialists due to increased medical specialisation. ICT is crucial for managing the increased communication resulting from this specialisation. Increased knowledge of an individual's genetic composition is also influencing the treatment.

The social factors that will affect development in 2012 are characterised by increased globalisation, an increased interest in one's own health and an increase in the elderly population. ICT helps people to stay in their own home and ICT is a prerequisite for globalisation.

The political development in Norway pertaining to healthcare is expressed as political decisions concerning healthcare services in the future. This is published in a national plan for healthcare services [1], a plan for how to provide care for, mainly, the elderly people [2], a report on how to provide continuity of care in healthcare services [3] as well as a report on the future use of ICT, in general, in Norway [4].

2.2. Technical Development

The technical development continuously produces new possibilities, but healthcare is often reluctant to use them. Much of the development that is expected to be further developed, are improved communication solutions and access of information on the internet.

Other technologies will allow things to be done in a new way, such as service oriented architecture (SOA). There is also an increased focus upon semantic interoperability compared to technical interoperability, which has been in focus for a long time. Computer assisted decisions (CAD) and system biology are also techniques that are expected to be of increased importance.

New technologies include gene-technology, intelligent information systems, semantic web and nanotechnology. Use of nanotechnology is expected to have significant effects within medicine, although probably not before 2012.

2.3. Development in Other Countries

We have performed an internet study on what is being developed in other countries. Most of the information available is not specifically related to the use of ICT, but to healthcare in general. In addition, a large portion of the information is also from countries where healthcare services may not be directly comparable to Norwegian (or North-European) healthcare.

The information available indicates a very coherent view on this immediate development. Common features include, among others:

- Focus upon EHR as well as patient summaries
- Web-access to EHR for authorised personnel, including the patient
- Increased focus on patient services, including self-treatment
- Centralised web access points for various services ("Portals")
- Definition of integrated treatment sequences
- Information exchange, including ePrescribing

However, we identified significant deviating views of opinion in two areas:

- Some parties intend to register structured information in the EHR while others propose to enter free text and to extract the information using data mining.
- Some parties plan to make patient information available from a central repository of all EHRs while others propose to make this information available from where it is originally registered.

3. Expected Main Features

3.1. Infrastructure

Use of ICT Equipment. ICT-equipment will become available everywhere and the physical size will continue to decrease. By the end of year 2012, a combination of various wireless communication methods will allow ICT-equipment units to be connected to a network everywhere: bedside in the hospital, at home, during travel, etc. making all needed information always available.

Better Coordination between Actors. Modern medicine requires the cooperation of several healthcare workers and medical specialists. ICT will not only provide the means to transfer the information between the actors, but will increasingly be used to keep track of the logistics in this cooperation involving all involved resources, such as personnel, facilities and equipment.

Improved Access to Patient Information. Authorised healthcare personnel involved in the treatment of a patient will have access to all relevant information related to the task the actual person is performing. This access will be provided everywhere, irrespective of the organisation of healthcare services and how the EHR is being stored. Increased availability of medical information reduces the need for storing duplicate copies by all parties involved in the treatment of a patient; each actor may access the information when needed from the original source or alternatively a central store.

Access Control. Increased availability to patient information requires a sophisticated access control system. All users have to be authorised by an official health authority. Access is granted only for selected parts of the total medical record related to the task the actual person is performing (dynamic role) and the patient's actual medical problem. Rules are being defined concerning which information has to be available for different healthcare workers in various parts of medical diagnostics and treatment.

Surveillance of Access by Patients. With increased access to patient information, the patient has to feel safe about the use of his/her own medical data. At any time, the

patients can access their own information on the web at "My Page"¹ and see which information has been accessed, when and by whom. As a side-effect, this will also allow the patients to follow the work that is being done by healthcare personnel concerning their own treatment.

Information Exchange. Information exchange using messaging (EDI) is still being used for operational tasks such as requests for services and reporting of status. However, the transfer of copies of documents from the patient's medical record is reduced since this information is available on demand from the original source.

Individually Adjusted Working Pattern. Various application systems² have been around for more than 30 years providing a more or less uniform user interface. These systems will be increasingly adjusted for the tasks to be performed by the actual user and will also provide guidance according to the user's wishes. Intelligent process support warns the user when abnormal situations are detected.

The user is also allowed to specify his/her user preferences concerning presentation and registration of information and which information elements this user wants to be available.

On-line Telemedicine. Traditional on-line telemedicine services like video-conferences are still being used to a limited extent for special purposes. However, the use of commonly available units like PCs and mobile phones with cameras are being used as an extension of the ordinary telephone consultation with GPs and specialists.

3.2. Electronic Healthcare Record

Increased Structuring of Information. Within the next 5 years, the information in the EHR is increasingly stored as structured information allowing application systems to use these information elements for more than just storage and presentation.

Dictation and voice recognition, used for inputting of large text blocks, are being replaced by intelligent, structured registration performed by the actual healthcare worker himself. The application system guides the data collection based upon the actual problem of the patient and asks for crucial, missing information. Combined with the features of individual user preferences, this way of entering information significantly increases the quality of the diagnostic process.

Intelligent business processes allow the application to propose relevant diagnoses based on all available information (including results from laboratories and other service departments as well as current best practice) and correspondingly warn the user when non-standard or potentially dangerous procedures are being initiated. This provides significantly increased safety for the patients.

Consistent Presentation of Information from Different Sources. Since information, to a larger extent, is being kept at its original location, the application system accesses these locations and presents the merged information as if it all should have been collected from the local storage.

Intelligent Presentation of Information. The storing of structured information allows for better presentation of information. The application system allows for easy navigation throughout the complete medical record of the patient only limited by the access granted for the actual purpose. The system is able to present chronological views of all contact episodes or to present the development of specific clinical

¹ "Min Side" in Norwegian.

² Such as PAS, HIS, RIS, etc.

parameters. The user may, at any time, select an item and ask for additional information. The information presented can be in graphical format or consist of pictures, videos, etc.

Overview of Ongoing and Scheduled Activities. The traditional patient summary of earlier episodes of disease and current treatment is supplemented with an overview of ongoing and scheduled activities in general also including such activities during hospital stays. This allows for coordination of planned activities and better information to the patient about ongoing treatment.

3.3. Medical Diagnostics and Treatment

Evidence Based Medicine. National guidelines for adequate medical treatment based upon evidence based medicine is published and made available to the vendors of medical software to incorporate in decision support, both inside hospitals as well as in primary care services.

Integrated Treatment Sequences. Complex medical problems require the assistance of various specialists in a predefined sequence. These integrated treatment sequences are also defined as a part of the national recommendations allowing the whole sequence of activities to be performed as a single procedure in a scheduled way, once started.

Individually Customised Treatment of Patients. Patients are treated more individually based upon, among others, their ethnic background. The genetic composition of patients is used to adjust the incidence of various diseases in diagnostic procedures, for evaluation of laboratory test results as well as for the dosage of medication.

Increased Patient Safety through Decision Support. The structured information in the EHR allows for continuous decision support in all patient processes. Decision makers, such as physicians and nurses, are guided in their work and warned if a decision is deviating from either the national recommendations or the patient's genetic composition, providing increased safety for the patient.

3.4. Use of Technology for Diagnostics and Treatment

Sensor-technology. Use of sensors will allow for better surveillance of patients and make it possible to start treatment as soon as an abnormal situation is detected.

Use of Simulators. The use of medical simulators for the training of medical personnel may be of special interest in Norway where personnel may not get the necessary training for complex situations, due to the limited population and geographic distribution. Such simulators are used for surgical procedures, medical emergency procedures, anaesthesia as well as other complex situations.

3.5. Patient Services

Self-treatment. The "modern" patient is much more interested in his/her own health and expects to participate in the medical decisions. By the end of year 2012, new application systems help the patient handle the enormous amount of information available on the internet and to evaluate the relevant information related to the patient's situation. The use of a personal electronic health record can be part of such tools or be

used separately. Personal electronic records ease the communication between the patient and involved health care personnel and participate to better treatment.

Electronic Communication between Patient and Healthcare. Secure e-mail is used for communication between the patient and assigned healthcare personnel. Simple questions may be solved without a physical consultation and pictures and videos may be attached. The ability to get in touch with qualified personnel, at any time, allows the patient to feel more comfortable so that the patient copes better with the situation.

Electronic communication is also used for scheduling appointments and for managing prescriptions through the web service "My Page". Here, the patient may also access his or her own medical record, view the patient summary, including on-going and scheduled activities, as well as get an overview of the usage of his/her medical record by others.

3.6. Auditing and Reporting

There is an increased interest for being able to monitor the effects of various health-related initiatives. The structuring of the EHR and the use of standardised definitions, allows managers to obtain national conclusions without having to establish a new national registry. The same feature allows for an improved and more up-to-date management of activities and costs, both locally, regionally and nationally.

4. Conclusions

The evolution is a continuous process and hopefully most of the expected new features will be operational by the end of 2012. However, wide-spread uptake of new technology is a fairly slow process and we, therefore, do not expect all features to be used by everybody by this date. In spite of this, we expect increased use of ICT in healthcare, by 2012, to provide, among others, increased service to patients and healthcare personnel, increased patient safety, better continuity of care as well as better diagnostics and treatment.

References

- [1] National Health Plan for Norway (2007-2010): http://www.regjeringen.no/upload/kilde/hod/anb/2006/0023/ddd/pdfv/296689-national_health_plan_for_norway_engelsk_del_1_02112006.pdf (accessed 25 February 2008).
- [2] Report No. 25 (2005 - 2006) to the Storting - Care Plan 2015 : <http://www.regjeringen.no/templates/RedaksjonellArtikkel.aspx?id=448249&epslanguage=EN-GB> (accessed 25 February 2008).
- [3] Wisløff-utvalget, NOU 2004:18 Helhet og plan i sosial- og helsetjenestene: <http://www.regjeringen.no/nb/dep/hod/dok/NOUer/2004/NOU-2004-18.html?id=150699> (in Norwegian only) (accessed 25 February 2008).
- [4] Report No. 17 (2006-2007) to the Storting - An Information Society for All: <http://www.regjeringen.no/en/dep/fad/Documents/Government-propositions-and-reports-/Reports-to-the-Storting-white-papers/20062007/Report-No-17-2006---2007-to-the-Storting.html?id=441497> (accessed 25 February 2008).

IT-based Information Management in Health Care Networks: the MedoCom Approach

Nils HELLRUNG^{a,1}, Nathalie GUSEW^a, Martin WILLKOMM^b, Reinhold HAUX^a

^a *Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig-
Institute of Technology, Germany*

^b *OFFIS Institute for Information Technology, Germany*

Abstract. Transinstitutional information systems support collaboration beyond the borders of single health care institutions. Information management in health care networks aims at establishing systematic information processing in order to contribute to the network's strategic goals. Health care networks can be characterized by complex and even conflicting goal systems. Our objective is to propose a framework, which describes organizational characteristics of health care networks. The framework is used to deduce effects of network characteristics on information management. Concluding, we present an architectural approach of a management platform for health care networks. The framework as well as the management platform are parts of our MedoCom approach.

Keywords. Management, Organization, Systems architecture

Introduction

The advancement of medical knowledge and medical technologies are leading to increasing functional differentiation in modern health systems. On the other hand, the coordination of health care activities between professionals, organizations and sectors is conceived as a key requirement for efficient and effective health systems [1], especially with regards to chronic diseases [2]. The concept of integrated care describes a principle of care organization, that encompasses continuity of care, shared care and seamless care [3]. Full institutional integration, i.e. the fusion of health care institutions to large hierarchical firms, has shown decisive drawbacks in providing integrated care, caused mainly by inflexibility and bureaucracy [4]. Therefore, in many countries efforts have emerged to achieve integrated care in health care networks, which consists of legally separated actors [1, 4-6].

Transinstitutional information systems, i.e. information systems that span the border of at least two legal independent health organizations, have been characterized

¹ Corresponding Author: Nils Hellrung, Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig-Institute of Technology, Germany, Muehlenfordtstr. 23, 38106 Braunschweig, Germany; E-mail: nils.hellrung@patri.de

as a key requirement for integrated and patient-centred care [3, 7, 8]. Given the relevance of information processing in health care and the complexity of organizational, technical and legal requirements, information systems should be managed systematically [9]. A central task of information management is the alignment of organizational goals and the respective information system architecture [10, 11]. In health care networks, which are characterized by mutual dependencies, complex goal systems and non-hierarchical governance forms [4, 12, 13], the coordination of collective network interests and interests of the individual network members with regards to information management remains an unsolved problem. Together with the Institute of Technology and Management of Technical University of Berlin and the OFFIS Institute for Information technology, we started the research project “IT-based management of integrated care networks” in September 2006. Part of the project is the development of an IT-based management platform that is named “MedoCom”.

The goal of this paper is to contribute to the development of transinstitutional information management, i.e. the management of transinstitutional information systems, in health care networks by presenting current results of our research. The following questions are addressed:

- Q1: How can health care networks be classified in a systematic description framework?
- Q2: How can network characteristics affect information management?
- Q3: How can transinstitutional information management be supported by IT?

1. Methods

Research question Q1 is answered by presenting a description framework for health care networks. The development was based on a literature review of publications from the field of health care and from other fields, such as economics and publications with a focus on organizational effects of transinstitutional information systems. Furthermore, we conducted empirical case studies in health care networks, which comprised modeling of central transinstitutional processes (using UML-activity diagrams) and the current state of the transinstitutional information systems (using the 3LGM² metamodel [14]). Research question Q2 is answered by analyzing the effects of network characteristics on decision making in health care networks. Schematic illustration of the architecture design is presented by a UML component diagram showing the most important modules and interfaces of the prospective architecture. Since the implementation of MedoCom is in a very early stage, Q3 is answered in the discussion and outlook sections.

2. The MedoCom description framework for health care networks

Health care networks can be defined as “autonomous units that have joined together to achieve a common purpose” [4]. We consider the autonomous units as individual persons, while the abstraction level can be varied to organizational units or whole organizations. Our framework aims at systematizing network characteristics that describe the relationships between the network members.

Transinstitutional relationships can be examined from various perspectives [15-17]. Similar to other approaches of organizational modeling, we therefore discriminate three interdependent levels of analysis: the management level, the process level and the infrastructure level (see **Table 1**).

Table 1: Levels of analysis in health care networks

	Focus of analysis (network)	Focus of analysis (network member)	Focus of analysis (relationships)
Management Level	structure of management system	management competencies	managerial dependencies
Process Level	process system	process elements assigned to network member	interweavement of process elements
Infrastructure Level	architecture of transinstitutional information systems	property rights regarding information resources	information dependencies

The management level focuses on the governance system of a health care network. The network members are described regarding their management competencies, i.e. their competencies of setting goals as well as planning, directing and controlling the network processes. The relationships between the members are conceived as managerial dependencies. The process level describes the collaborative process chain of the health care network. Network members are analyzed with regards to their assigned process elements. Relationships are interpreted as processual dependencies. At the infrastructure level, the architecture of the transinstitutional information system is observed. Network members are examined regarding their rights to use, to change or to control (i.e. their property rights) information resources. According to [14, 18], we consider information resources, as application components or physical data processing

Table 2: Centrality and intensity as characteristics of health care networks

	Centrality		Intensity	
Management Level	allocation of managerial authority		extent of managerial influence	
	low: managerial authority is equally shared	high: managerial authority is concentrated on one or few members	low: managerial influence on collaboration is weak	high: managerial influence on collaboration is high
Process Level	allocation of processes		quantity of process execution	
	low: processes are assigned equally	high: processes are mainly assigned to one or few members	low: processes are executed infrequently	high: processes are executed frequently
Infrastructure Level	allocation of information resources		quantity of resource usage	
	low: information resources are controlled equally	high: information resources are controlled by one or few network members	low: information resources are used infrequently	high: information resources are used frequently

components that support the execution of activities. The relationships between the network members are conceived as dependencies that result from varying allocations of property rights. Using our framework, we developed characteristics that help us to specify different forms of health care networks. Two of these characteristics are centrality and intensity (see **Table 2**). Centrality is a network characteristic that describes the structure of a network regarding the allocation of power between the network members. Intensity describes the extent of actual network activities.

3. Effects of networks characteristics on information management

Transinstitutional information management can be defined as the sum of all planning, directing and monitoring activities with regards to a transinstitutional information system. In health care networks the goal of information management is systematic information processing that contribute to the network's strategic goals [9, 19].

The effects of decentralization on information management in health care networks can be summarized as follows: due to the legal autonomy of the network members, no hierarchical structures exist. Hence, mechanisms for the coordination of multiple goal systems have to be developed. This includes not only activities during the planning phase but also during the directing and monitoring phase. Second, the extent of decision implementation is self-governed by each network member. With decreasing network intensity, i.e. with decreasing relevance of the network, the willingness of implementing network decisions will decrease as well. Therefore, the relevance of the network for the individual member as well as the relevance of the member for the network has to be evaluated. Third, due to the voluntariness of network membership, the network composition will change dynamically. Therefore, information management must be adaptable to the current network situation. In figure 1, management decisions in centralized and decentralized networks are compared.

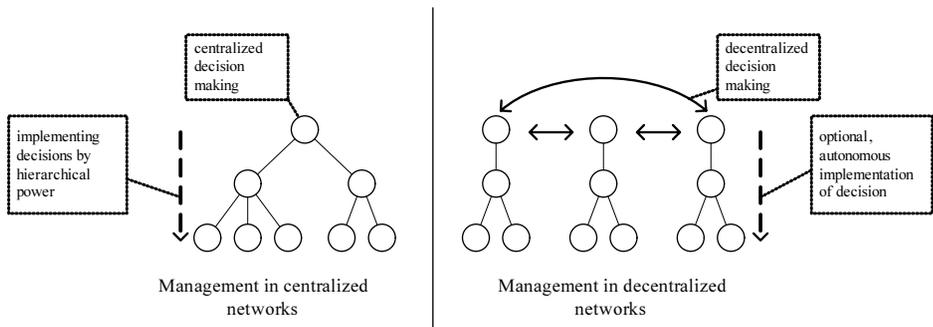


Figure 1: Management decisions in centralized and decentralized networks

4. Discussion

Our empirical research showed that the current structures of health care networks are often very informal. Although most of the collaboration problems are connected to information problems, the need for systematic information management is not always perceived. The preceding framework suggests a methodology for systematically

describing health care networks considering network characteristics at different examination levels. It is our starting point for the development of MedoCom, an IT-based platform for cooperative management in health care networks.

The next development step of MedoCom will be the extension as well as the formalization of the description framework. This includes measurable and comparable parameters for describing networks, e.g. centrality and intensity. Since organizational modeling has been subject of several research approaches before, we will adapt suitable modeling approaches, e.g. Business Process Modeling Notation [20] for process modeling and 3LGM² [14] for architectural modeling. We are confident that our approach of characterizing health care networks will contribute to the development of transinstitutional information management.

However, MedoCom has yet to be introduced and evaluated in real health care networks. A prototypical implementation of our concept is planned for spring in 2008. In the outlook section, our architectural concept of MedoCom is presented.

5. Outlook

Figure 2 depicts the architectural approach of MedoCom and shows the component diagram of MedoCom architecture design consisting of primary modules and communication interfaces of the application. Currently, the network model consists of a static network view and the business process view. Since the network model will serve as the base for management decisions, it is necessary to update it constantly. Therefore, the model will be linked to operative applications based on web services (eServices). At the present time, service oriented architecture (SOA) approach is most promising and broadly approved in order to address problems related to the integration of heterogeneous applications in distributed environments [21].

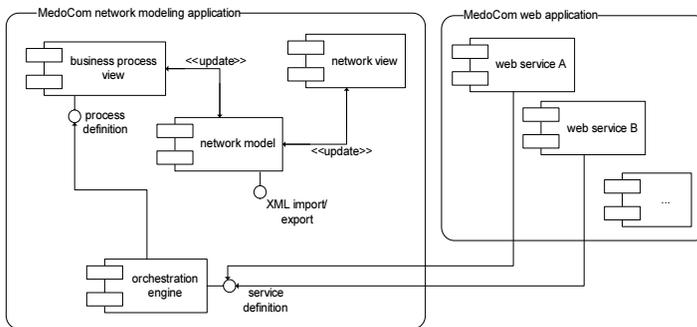


Figure 2: Architectural approach of MedoCom

Thus, the modelled activities will be coupled to an executive component and could be requested and accessed by the users over the internet or web-based transinstitutional information systems. Running the web applications by network participants would mean the initiation of activities for one particular user, who supplies new data to the model by passing the process steps. The additional information would be fed back to the application and auto-generate a new adapted network model.

References

1. Wijngaarden JDH, de Bont AA, and Huijsman R. Learning to cross boundaries: The integration of a health network to deliver seamless care. *Health Policy* 2006. 79(2-3): p. 203-213.
2. Anderson G and Knickman JR: Changing The Chronic Care System To Meet People's Needs. *Health Affairs* 2001. 20(6): p. 146-160.
3. Kuhn KA, Giuse DA, Lapao L, Wurst SHR: Expanding the Scope of Health Information Systems. *Yearbook of Medical Informatics*, 2006. 2006: p. 43-52.
4. Alexander JA , Lee SY, Bazzoli GJ, Governance Forms in Health Systems and Health Networks. *Health Care Management Review*, 2003. 28(3): p. 228-242.
5. Greß S, Focke A, Hessel F, Wasem J: Financial incentives for disease management programmes and integrated care in German social health insurance. *Health Policy*, 2006. 78(2-3): p. 295-305.
6. Mur-Veeman I, Hardy B, Steenbergen M, Wistow G: Development of integrated care in England and The Netherlands: managing across public-private boundaries. *Health Policy*, 2003. 65(227-241).
7. Hellesø R and Lorensen M: Inter-organizational continuity of care and the electronic patient record: A concept development. *International Journal of Nursing Studies*, 2005. 42: p. 807-822.
8. Mykkänen J, Korpela M, Ripatti S, Rannanheimo J, Sorri J: Local, Regional and National Interoperability in Hospital-level Systems Architecture. *Methods Inf Med*, 2007. 46(4): p. 470-475.
9. Haux R, Winter A, Ammenwerth E, Brigl B: *Strategic Information Management in Hospitals - An introduction to Hospital Information Systems*. 2004, New York, Berlin, Heidelberg: Springer-Verlag.
10. Wainwright D and Waring T: Three domains for implementing integrated information systems: redressing the balance between technology, strategic and organisational analysis. *International Journal of Information Management*, 2004. 24(4): p. 329.
11. Winter A., Ammenwerth E, Bott OJ, Brigl B, Buchauer A, Gräber S, Grant A, Häber A, Hasselbring W, Haux R, Heinrich A, Janssen H, Kock I, Penger OS, Prokosch HU, Terstappen A: Strategic information management plans: the basis for systematic information management in hospitals. *International Journal of Medical Informatics*, 2001. 64: p. 99-109.
12. Janus K. and Amelung VE: Integrated Health Care Delivery Based on Transaction Cost Economics: Experiences from California and Cross-National Implications. *Advances in Health Care Management*, 2005. 5: p. 117-156.
13. Meijboom B, de Haan J, and Verheyen P: Networks for integrated care provision: an economic approach based on opportunism and trust. *Health Policy*, 2004. 69: p. 33-43.
14. Winter A, Brigl B, Wendt T: Modeling Hospital Information Systems (Part 1): The Revised Three-Layer Graph-Based Meta Model 3LGM2. *Methods Inf Med*, 2003(5): p. 544-551.
15. Kodner DL and Spreeuwenberg C, Integrated care: meaning, logic, applications, and implications - a discussion paper. *International Journal of Integrated Care*, 2002. 2.
16. Luke RD: Taxonomy of Health Networks and Systems: A Reassessment. *Health Services Research*, 2006. 41(3): p. 618-628.
17. Dubbs NL, Bazzoli GJ, Shortell SM, Kralovec P: Reexamining Organizational Configurations: An Update, Validation, and Expansion of the Taxonomy of Health Networks and Systems. *Health Services Research*, 2004. 39(1): p. 207-220.
18. Brigl B, Wendt T, and Winter A: Modeling interdependencies between information processes and communication paths in hospitals. *Methods Inf Med*, 2006. 45(2): p. 216-224.
19. Haux R: Health information systems - past, present, future. *International Journal of Medical Informatics*, 2006. 75(3-4): p. 268-281.
20. Object Management Group: Business Process Modeling Notation. 2004 [cited 2007 November 11th]; Available from: <http://www.bpmn.org/Documents/BPMN%20V1-0%20May%2003%202004.pdf>.
21. Mukhi N, Konuru R and Curbera F: Cooperative middleware specialization for service oriented architectures. in: *Proceedings of WWW 2004*. 2004.

The Role of Patients and their Health Cards in Integrated eHealth Environments

Claudia HILDEBRAND ^{a 1}, Peter PHAROW ^b and Bernd BLOBEL ^b

^a *Helmholtz Zentrum München - German Research Center for Environmental Health,
Neuherberg, Germany*

^b *eHealth Competence Center, Regensburg University Hospital, Regensburg, Germany*

Abstract. Communication and co-operation processes in healthcare and welfare require the involvement of all parties involved, including health professionals as well as patients. Generally, professionals can and will easily communicate via trusted health networks. To enforce both communication and co-operation between professionals and patients and to guarantee the required degree of involvement of patients in shared care environments, smart cards are widely used. They serve as person identifiers on the one hand and as security token on the other hand. Acting as storage media and portable application systems, patient data cards enable patient-controlled exchange and the use of personal health data for specific purposes such as prescription and disease management. Additionally, patient status data such as the emergency data or the immunisation record may be stored in and communicated by patient data cards.

Keywords. Health Network, Health Cards, Security, Infrastructure, Patient Data Cards, Policy, Electronic Health Record System, Patient Integration

Introduction

Knowledge accumulated in the recent decades proved that information that flows within a domain and its transfer between domains is the very basis of human progress. Professional knowledge becomes more valuable when shared with other professionals. This is also true for healthcare and welfare. In order to increase quality and efficiency of healthcare, health systems in developed countries throughout the world tend to move towards distributed collaborative and coordinated care of patients in the sense of shared care. Additionally, health care administration and management focus more and more on preventive care which the citizen himself/herself is responsible for. Regardless whether integration of information is considered for domains like bio-informatics, biomedical engineering, applied and clinical informatics, or for information systems and informatics education, it is always collaboration and exchange of knowledge that counts [1].

¹ Corresponding Author: Claudia Hildebrand, Helmholtz Zentrum München - German Research Center for Environmental Health (GmbH), Institute of Biological and Medical Imaging, Ingolstaedter Landstr. 1, 85764 Neuherberg, Germany

Patient empowerment and patient involvement in the healthcare and welfare processes are important pre-requisites for achieving the positive outcomes hoped for. Health cards can play an important role. They allow developing an integrated eHealth environment by integrating medical informatics, health informatics, patient integrity, and patient integration. Keywords to foster these processes are card, security token, and card technology. The role of the card in the process of patient integration and the security functionality for data integrity and patient integrity and the problems concerning access rights, and related data protection and privacy aspects are also vital [2].

1. Patient Empowerment and Patient Involvement

From the aforementioned point of view, healthcare systems will increasingly profit from engaged citizens while engaged patients will have better outcomes. New health policies, increasing patient needs and decreasing health budgets necessitate empowerment of the patient. The involvement has to take place at different levels within the healthcare process at the level of care as well as at the political level [3]. At the level of care, the interest in citizen engagement is presently changing from past practices where the patient's involvement meant that the patient had to go along with the physician's advice- to a collaborative partnership where the citizen is considered a partner within the medical workflow. At political level patient group representatives are increasingly considered valuable allies [4], [5].

Patient empowerment means that citizens have to take on responsibility. Citizens have to realise that healthcare is a concern of theirs even when they are not in need of it. This means they have to understand the system. Only if they understand the system and know how to influence the healthcare process, will the citizens build up the necessary confidence in the new system and will be ready to accept any change of behaviour that might be required. Health care providers and politicians, on the other hand, have to realise that healthcare systems and services can only be improved with actively participating citizens.

A dedicated patient health card (PHC) as a means to strengthen the role of the citizen has multiple functionalities [6]. It can be used for purely administrative reasons, but also support the communication between care providers and improve data security. In either case the patient has to support this process – at minimum – by presenting the card whenever required. This means that the patient has to be involved but is in the same time empowered. However, this kind of involvement asks for a certain level of process understanding (awareness, confidence, acceptance), too [7]. How can a card support this?

2. Patient Empowerment by Standardization in the Patient Health Cards Domain

Open, interoperable and scalable solutions for empowering and involving citizens and patients worldwide especially if they are to include respective cross-border activities must be based on international standards [8]. Otherwise a French health professional, for example, will not be able to read the health card of his/her German patient and vice versa.

Regarding the aforementioned requirements for interoperable patient data cards, a series of standards have been specified at international level by ISO TC 215 “Health Informatics”. After having identified both needs and requirements for such cards, the standardization experts defined a framework of card-related specifications of medical and administrative content. ISO TC 215 does not intend to standardise the card-related technology itself but only the health-related structures of such types of cards.

Person-related data carried on a data card can be categorised into three types: identification data (of the device itself and of the individual the data it carries relates to), administrative data related to the card owner, and the respective medical (clinical) data. It is important to realise that a given healthcare data card "de facto" has to contain device data and identification data and can contain, in addition, administrative and clinical data (see figure 1 below). Furthermore, patient data cards may support the collaboration with network-based systems. For that purpose, any type of link information was specified. Person-related cards can also enable the use of established security infrastructure services.

Developed under the rules of the so-called “Vienna Agreement” for allowing the European Standardization Body CEN to easily adopt existing ISO standards, ISO TC 215 WG 5 “Health Cards” has started developing standard 21549 “Health Informatics – Patient health card data” [9]. This new standard has started replacing the former European pre-standard ENV 12018 adopted by CEN in 1995. ISO 21549 consists of eight parts describing structures and roles within a common framework.

3. International Standard ISO 21549

A data card essentially provides specific answers to definite queries whilst at the same time there is a need to optimise the use of memory by avoiding redundancies. Figure 1 shows the overall structure for patient health card data according to the described 8 parts of ISO 21549 [9] using a UML Class Diagram.

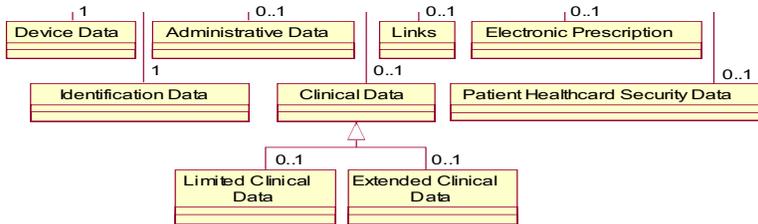


Figure 1. ISO 21549 Data Set

3.1 Part 1 – General Structure

The first part is more or less an overview on what needs to be tackled by such a series of standards (see figure 1). Part 1 is the introduction to the multi-part standard that defines data structures kept on patient health cards compliant with the physical dimensions of ID-1 cards as defined by ISO/IEC 7816. Therefore, this part does not

apply to multi-application cards. It defines a general structure for the different types of data, which are defined in the other parts of the standard, using UML notation. ISO 21549 Part 1 has been in place since May 2004. Several countries are basing their national health card implementation strategies on this standard.

3.2 Part 2 – Common Objects

Part 2 is dedicated to data items and structures that need to be present on each and every single health card to be addressed and accessed. Thus, part 2 establishes a common framework for the content and the structure of common objects used to construct other data-object data held on patient healthcare data cards, or references thereof. This part specifies the basic structure of the data but does not specify or mandate particular data-sets for storage on devices. The latter will be done in the other parts. Part 2 of ISO 21549 has been in place as a standard since May 2004.

3.3 Part 3 – Limited Clinical Data

Part 3 defines and describes the limited clinical data objects used in, or referenced by, patient-held health data cards using UML, plain text and abstract syntax notation (ASN.1). Part 3 specifies the basic structure of the data contained within the data object “Limited Clinical Data” but does not specify or mandate particular datasets for storage on devices. In particular, the data contained within the data objects in “Limited Clinical Data” are intended to aid the delivery of emergency care, but are by themselves neither intended, nor suitable, for the provision of all information required. ISO 21549 part 3 has been in place since May 2004.

3.4 Part 4 – Extended Clinical Data

Part 4 specifies the basic structure of the data contained within the data object “Extended Clinical Data”, but does not specify or mandate particular data-sets for storage on devices. In order to facilitate interoperability, whenever an application is built for use in the healthcare domain in compliance with this standard, data items required for that application shall be drawn from a defined list of objects. These shall then be used in conjunction with other data defined in other parts of this standard.

3.5 Part 5 – Identification Data

Part 5 establishes a common framework for the content and the structure of “Identification Data” of the device-holder held on healthcare data cards. It specifies the basic structure of the data, but does not specify particular data-sets for storage on devices. Its structures can accommodate suitable data objects specified elsewhere

- Security functions and related services which are likely to be specified by users for data cards depending on their specific application, for example confidentiality protection, data integrity protection, and authentication of persons and devices related to these functions;
- Initialization and issuing processes (which starts the operating lifetime of any individual data card, and by which the data card is prepared for the data to be subsequently communicated to);

- Access control services which depend on active use of some data card classes such as microprocessor cards.

3.6 Part 6 – Administrative Data

Part 6 specifies the basic structure of the data contained within the data object “Administrative Data”, but does not specify or mandate particular datasets for storage on devices. In order to facilitate interoperability, whenever an application is built for use in the healthcare domain in compliance with this standard, data items required for that application shall be drawn from an existing list of objects (some of which are extensible). These shall then be used in conjunction with other data defined in other parts of ISO 21549. For delimiting the administrative data set of this standard from the identification data set of ISO 21549-5 the administrative data set shall contain health insurance data only.

3.7 Part 7 – Electronic Prescriptions

Part 7 specifies the basic structure of the data contained within the medication data object. It defines the structure for electronic prescriptions. This is on top of the priority list of most European countries as this use case seems to be the most promising concerning acceptance and cost savings [10].

3.8 Part 8 – Links

Part 8 is of specific importance for the future use of health cards as this part intends to bridge card and network. Cards will not be able to bear complete information (plain text, structured information, images, other multi-media information, etc.) regarding a patient’s health status. Networks will easily communicate this information within a very short time. Cards will therefore act as keys to this information. To allow for all these functions, links and other pointers need to be stored on the card (URL, URN, tickets, etc.). Part 8 of ISO 21549 will define and standardise this kind of structure and both EFMI WG “CARDS” and EFMI WG “Electronic Health Record” will play an important role in these processes [11]. Draft versions of Part 8 have been extensively discussed during 2007.

4 Conclusions

Shared care solutions all over the world have to be based on trustworthy communication and application security services. Patient Identification Cards (PIC), Patient Data Cards (PDC), and Health Professional Cards (HPC) will play an important role; either as personal ID token, as professional ID token, or as health data carriers. Cards have an impact on the related security infrastructure, the certification of processes, on process interoperability (workflow), and on the certification of state and relations of principals in longer terms. This is especially true for the upcoming developments on Electronic Health Record (EHR) architectures, their requirements, their design, their policy details, and their instantiation and implementation strategies.

Citizen involvement and patient empowerment are based on awareness, confidence, and acceptance. An adequate level of awareness can be achieved by the respective level of information, education and training measures. This will lead to confidence with regard to shared care, the new healthcare paradigms but also the health card. With the regular use and based on the patients' respective behaviour, the level of acceptance is going to grow leading to an even higher level of awareness at the end of the day. Additional applications like the electronic signature, as a first step towards a real multi-application card, can foster these processes.

The more information citizens and patients do have regarding different procedures and processes in healthcare and welfare, the more they are able to significantly play their dedicated role within this partnership. Cards can and will contribute by allowing citizens and patients to get controlled access to administrative and medical data stored either on a card or in the network but also to determine who else shall have access to this data.

Acknowledgement

The authors are in debt to the European Commission for the funding of several European research projects as well as all other partners and organisations for their support and co-operation.

References

- [1] B. Blobel: Analysis, Design and Implementation of Secure and Interoperable Distributed Health Information Systems. Series "Studies in Health Technology and Informatics" Vol. 89. IOS Press, Amsterdam 2002.
- [2] B. Blobel, P. Pharow (Eds.): Advanced Health Telematics and Telemedicine. The Magdeburg Expert Summit Textbook, pp. 21-28. Series "Studies in Health Technology and Informatics" Vol. 96. IOS Press, Amsterdam 2003
- [3] International Alliance of Patients' Organizations (IAPO), <http://www.patientsorganizations.org> (last accessed February 14th, 2008)
- [4] IAPO Policy Statement on Patient Involvement, <http://www.patientsorganizations.org/showarticle.pl?id=590&n=962> (last accessed February 14th, 2008)
- [5] World Health Organization. World Alliance for Patients Safety. Geneva, 2005.
- [6] H. Demski, C. Hildebrand, R. Engelbrecht, C. Birkmann: DIABCARD, the portable patient record – pursuing routine use. In: Blobel B. et al. (Eds.) Contributions of Medical Informatics to Health. Berlin, 2004.
- [7] P. Pharow, B. Blobel: Security Infrastructure Requirements for Electronic Health Cards Communication. In: Engelbrecht R., Geissbuhler A. (Eds.): Connecting Medical Informatics and Bio-Informatics. Proceedings of MIE 2005. Series "Studies in Health Technology and Informatics" Vol. 116. IOS Press, Amsterdam 2005.
- [8] C. Hildebrand, H. Demski et al: BioHealth – The Need for Security and Identity Management Standards in eHealth. In: L. Bos, L. Roa, K. Yogesan, et al. (Eds.): Medical and Care Compunetics 3, pp.327-336. Series Studies in Health Technology and Informatics, Vol. 121. IOS Press, Amsterdam, 2006.
- [9] ISO 21549 "Health Informatics – Patient health card data", 2004. <http://www.iso.org> (last accessed January 16th, 2008)
- [10] S.T. McMullin, T.P. Lonergan, C.S. Rynearson: Twelve-month drug cost savings related to use of an electronic prescribing system with integrated decision support in primary care. *J Manag Care Pharm.* 2005 May; 11(4):353-5, 356-7.
- [11] ISO 21549 "Health Informatics – Patient health card data, part 8 Links", 2006 (working draft)

A Deployment and Research Roadmap for Semantic Interoperability: the EU SemanticHEALTH project

Pierre LEWALLE^c Jean M. RODRIGUES^{a,1}, Pieter ZANSTRA^b, Bedirhan USTUN^c,
Dipak KALRA^d, Gyorgy SURJAN^c, Alan RECTOR^f, Veli STROETMANN^g, Martti
VIRTANEN^h

^a *University of Saint Etienne France*

^b *WHO-FIC Collaborating Centre in the Netherlands*

^c *WHO Classifications and Terminology Geneva Switzerland*

^d *University College London United Kingdom*

^e *ESKI Budapest Hungary*

^f *University of Manchester, United Kingdom*

^g *Empirica GmbH Germany*

^h *WHO-FIC Collaborating Centre in the Nordic countries, Sweden*

Abstract: The purpose of this EU funded project is to describe a short and medium term Research and Deployment Roadmap for Semantic Interoperability in e-health. It started by defining 4 levels and 3 dimensions for Semantic Interoperability. The vision is to reconcile the needs for the direct patient care safety, biomedical and clinical research and for public health by the reuse of direct care data.: from gene to individuals and populations. The methodology is presented and preliminary results and milestones for the short and the long term are set. We conclude by statements on the main characteristics and needs of the roadmap to sustain better health for individual and populations in the changing EU health care systems..

Keywords: Semantic Interoperability; Public Health; Electronic Health Record; Ontology; Multilingualism;

Introduction

Semantic Interoperability (SIOp) is widely considered a key ingredient for meeting present and future challenges of health systems such as ageing populations and increasing medical costs. The open challenges for SIOp result both from the clinical settings perspective of seamless care provision across all elements of the health care value system (like patient safety, multiple carers locations of health care delivery), from the research perspective of integrating life sciences and clinical information and

¹ Corresponding Author : JM Rodrigues, CHU de St Etienne, SSPIM/CIM42, Chemin de la Marandière , 42 270 Saint Priest en Jarez , France, rodrigues@univ-st-etienne.fr

knowledge, and from the public health perspective of policy monitoring based on data from direct patient care settings. Chapter 2 of this paper describes the chosen methodology to create the roadmap, chapter 3 gives outlines preliminary results with an overview of the various levels of interoperability and chapter 4 briefly outlines the initial roadmap.

1. Methodology of the Roadmap

The roadmap follows the requirements of the technology roadmap program as presented in [1] [2] [3]. It is designed to uncover the key issues that hamper efficacious SIOp in the health care system. For some problems the solutions are obvious and require adequate planning for implementation. Others do need more (basic) research to elicit a proper solution. This means different time scales for achieving different levels of SIOp. The roadmap project produces medium- and long term recommendations for research and deployment. We have defined five types of realistic use cases. A roadmap has three stages: preliminary activities, development, and follow-up [1][2][3].

2. Preliminary Activities

In the preliminary stage, the project (a) has ensured that the essential conditions for start-up were met, (b) has found the leadership and participants, and (c) has defined the scope and boundaries for the roadmap.

The literature and practice in SIOp has been reviewed. The outcome of that exercise is presented in Deliverables D1.1, [4] D1.2 [5], D2.1 [6], D3.1 [9], D5.1[7], D7.0[8] and RIDE deliverable D4.1.1 [10]. It provides the baseline (common conceptual framework, inventory of existing initiatives, technology and socio economic issues, public health issues) for the roadmap development. Stakeholder buy-in was sought to ensure a wide participation of key interested parties in the road mapping effort. This was done in several workshops involving Member States representatives (e-Health Stakeholders Group, the Continua Alliance, e-Health Interoperability Expert). Some have been dedicated workshops organised by SemanticHEALTH, and others have been sessions run in parallel with larger scale or focussed events, in order to maximise the range of inputs obtained.

3. Development of the Roadmap

The development of the roadmap rests on a careful consideration of enabling technology. The analytical dimension covers both technical and socio-economic aspects of SIOp [4][5][6][8]. The application dimension focuses on 3 specific application fields: The direct patient care with Electronic Health Record (EHR), the public health electronic record and the research dimension on ontology and multilingualism. The different steps have been processed and the final recommendations will be available in the middle of 2008: a common conceptual framework based on the vision statements of the purpose and goals of the defined SIOp in Deliverable D1.1 [4], an inventory of the national and international e-health initiatives and assessment of their SIOp performance

in Deliverable D1.2[5], the enabling factors: legal, political agreements, financial, social, organisational in Deliverable D3.1[9], technologies and products with their critical and secondary specifications to be developed by industry in Deliverable D2.1 [6], definition of 5 major categories of use cases (patient care, public health, research and translational medicine, support for diverse markets and cost drivers) in Deliverable D7.1[11], a scenario-based planning.(developed by the e-Health Interoperability Expert Group for implementing Large Scale Pilots: patient summaries and e-prescribing[5], identifying the gaps against real situations in Deliverables D4.1[12], D5.1[7] and D6.1[13] and the major areas of technology and enabling factors to be mobilised and their milestones .

4. Follow-up

The group of experts who developed and drafted the technology roadmap is relatively small. To enable a broad acceptance of the roadmap and to ensure future action, critical assessments - validated and accepted by a large audience including clinical bodies - is of utmost importance. To avoid the process coming to a halt over the course of the roadmap definition, all participants agreed to develop an action plan advocating investment decisions and setting out the means and time lines for implementation. Finally, since both the needs and the technologies are constantly and rapidly evolving, the implementation plan includes provisions for the periodic review and update of the initial roadmap. This activity was developed with the involvement of the RAG (Roadmap Assessment Group).

The roadmap document has been distributed to a large group of stakeholders representatives, who will thoroughly evaluate it and validate it or possibly suggest modifications. These reviewers will be asked to address specific questions such as the following:

1. If the recommended alternatives are developed, will the targets be met?
2. Are the technology alternatives reasonable?
3. Were any important alternatives or enabling factors missed?
4. Is the roadmap clear and understandable?
5. Are the recommendations feasible?
6. Can the recommended actions be completed in the required time frames?

5. Preliminary Results

As a specific support action, SemanticHealth is based on real existing implementations, or implementations planned for the near future. It distinguishes 4 levels of interoperability, 2 of them allow for semantic interoperability. To explain and distinguish the 4 different levels, consider the following scenario: 56 year old Pádraig recently moved from Ireland to Spain to take up his new job in a multinational IT company. A few weeks after arriving, he falls ill, consults his local (Spanish) GP and his being transferred to the next hospital for further tests.

- **Level 0** (no interoperability at all) – Pádraig has to undergo a full set of lengthy investigations to find out the cause of his severe pain. Unfortunately, results from the local GP as well as from his Irish GP are not available at the point of care within the hospital due to e.g. the missing technical equipment.

- **Level 1** (technical and syntactical interoperability) – Pádraigs doctor in the hospital is able to receive electronic documents that were released from the Irish GP as well as his local GP upon request. Widely available applications supporting syntactical interoperability (such as web browsers and email clients) allow the download and provide immediate access. Unfortunately, none of the available doctors in the hospital is able to translate the Irish document, and only human intervention allows interpreting the information submitted by the local GP for adding into the hospitals information system.
- **Level 2** (partial semantic interoperability) - The hospital doctor is able to securely access via the Internet parts of Pádraig Electronic Health Record released by his Irish GP as well as the local GP that he visited just hours earlier. Although both documents contain mostly free text, fragments of high importance (such as demographics, allergies and medical history) are encoded using international coding schemes, which the hospital information system can automatically detect, interpret and meaningfully present to the attending physician. Within the notion of partial SIOp, the term degree is being used to identify the degree of SIOp, e.g. 40% or 90%. A degree of 0% partial SIOp describes a situation where no SIOp is available whereas 100% identifies full SIOp. Two subcategories have been identified to distinguish between unidirectional semantic interoperability (level 2a) and semantic interoperability of meaningful fragments (level 2b).
- **Level 3** (full semantic interoperability, co-operability) - In this ideal situation and after thorough authentication took place, the Spanish hospital information system is able to automatically access, interpret and present all necessary medical information about Pádraig to the physician at the point of care. Neither language nor technological differences prevent the system to seamlessly integrate the received information into the local record and provide a complete picture of the Patients health as if it would have been collected locally. Further, the anonymised data feeds directly into the tools of public health authorities and researchers.

6. Initial Roadmap

The roadmap shall show the road for the deployment of the different solutions to the requirements of the European e-Health policies on: the interoperable structured patient summaries, intermediate interoperable EHRs, fully interoperable EHRs and regional public health data bases from interoperable EHRs.

The roadmap shall use the technical standards when they are available: EHR international standard (ISO), common nomenclature and agreed equivalence among medical products, the setting of an international infrastructure to support interoperation of ontology-based multilingual terminologies and the definition of a public health metadata from an EHR.

Finally the roadmap shall comply with legal rules when they are available: liability and responsibilities of the actors involved in the exchange of EHRs, legal transparency and liability concerning the e-Prescription, data protection compliance to patient privacy. The overall representation of the preliminary roadmap is illustrated below.

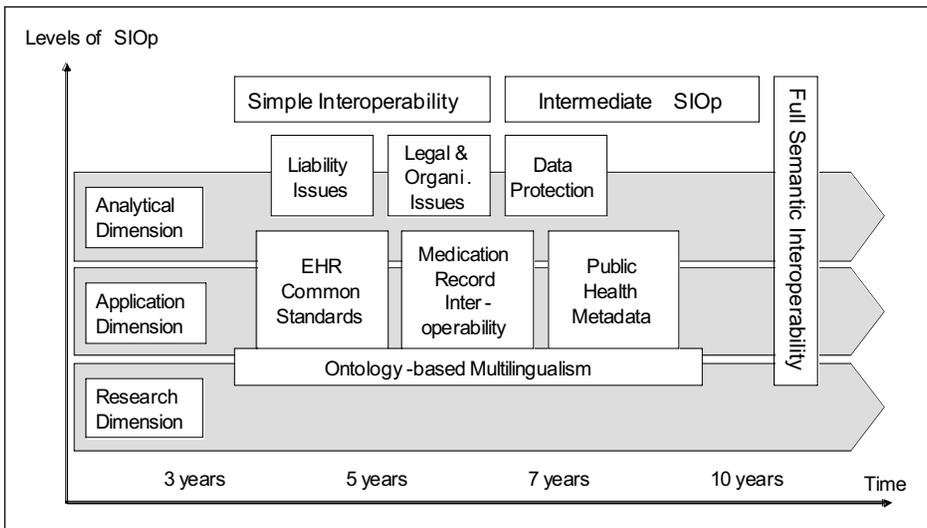


Fig.1 The overall representation of the preliminary roadmap

On a 10 years time line we represent the three levels of SIOp corresponding to the three deployment required by the eHealth policies. The three dimensions (analytical, application and research) are represented by the three horizontal stripes.

On the short term (5 years) the level 1 SIOp can be achieved with the implementation of patient summaries and ePrescription subject to compliance with legal rules (responsibilities of actors and legal organisation of e-prescription), and to the approval by ISO of the EHR standard and to an agreement on a common nomenclature of medical products.

On the medium term (5 to 7 years), if Level 2 SIOp is to be achieved it is necessary that legal rule on data protection compliance to patient privacy within an interchangeable EHR environment and a definition of the public health metadata from EHR are set.

Finally on a 10 years perspective the Level 3 SIOp can be achieved when an international infrastructure based on an ontology repository supporting multilingual terminologies and on repositories of structured EHRs is set.

7. Conclusion

The SemanticHEALTH roadmap is based on three statements, an environment framework and four types of actions:

1. SIOp is necessary for (a) translational medicine, (b) speed, and (c) coping with knowledge magnitude, not a goal of itself. The efforts needed to promote SIOp shall be aligned on scalable goals and realistic time schedule before showing the positive impact. There is a significant net social gain from SIOp, i.e. the benefits exceed the costs for quality, access, cost effectiveness paradigm, communication in healthcare between all actors including the citizen, care transfer and safety.
2. An environment framework is needed to initiate a paradigm shift from technical products to sustainable processes and structures (including legal

requirements, economic analysis and policy management), to design a dynamic, language-independent, sustainable reference repository of terminology, a cost benefit model allowing to decide the level of desirable SIOp and of scalability and a step-by-step test of conformance to SIOp.

3. Four types of actions shall be planned: (a) adoption of existing solutions, (b) wide-scale evaluations, (c) investment in development and (d) further research.

We are convinced that the proposed roadmap offers a viable option for better health in the currently changing health care systems.

Acknowledgements

Semantic Interoperability Deployment and Research Roadmap (SemanticHEALTH) is a Specific Supportive Action funded by the EU commission under FP7:IST-27328-SSA.

References

- [1] WIKIPEDIA Technology roadmap. http://en.wikipedia.org/wiki/Technology_roadmap
- [2] Galvin R; Science roadmaps in Science, vol. 280, p. 803, May 8, 1998
- [3] Garcia M L and Bray O H. Fundamentals of technology roadmapping. Sandia Nat. Labs., Albuquerque, NM, SAND97-0665, March 1998
- [4] SemanticHEALTH Deliverable D1.1. Conceptual Framework for e-Health Interoperability. European Union Commission; Brussels 2006.
- [5] SemanticHEALTH Deliverable D1.2. Interoperability: inventory of key relevant member states and international experiences. European Union Commission; Brussels 2007.
- [6] SemanticHEALTH Deliverable D2.1. Towards a semantic interoperability roadmap: technological issues. European Union Commission; Brussels 2007.
- [7] SemanticHEALTH Deliverable D5.1. Towards a semantic interoperability roadmap: public health and secondary uses. European Union Commission; Brussels 2007.
- [8] SemanticHEALTH Deliverable D7.0. Methodology for a roadmap; European Union Commission; Brussels 2007.
- [9] SemanticHEALTH Deliverable D3.1. Comparative analysis and socio-economic recommendations for improving semantic interoperability. European Union Commission; Brussels 2007.
- [10] RIDE Deliverable D4.1.1. European Union Commission; Brussels 2006.
- [11] SemanticHEALTH Deliverable D7.1. Semantic interoperability deployment and research roadmap European Union Commission; Brussels 2008.
- [12] SemanticHEALTH Deliverable D4.1. Barriers, approaches and research priorities for semantic interoperability in support of of clinical care delivery European Union Commission; Brussels 2008.
- [13] SemanticHEALTH Deliverable D6.1. Barriers, approaches and research priorities for integrating biomedical ontologies European Union Commission; Brussels 2008.

DebugIT for Patient Safety – Improving the Treatment with Antibiotics through Multimedia Data Mining of Heterogeneous Clinical Data

Christian LOVIS^{a,1}, Dirk COLAERT^b, and Veli N. STROETMANN^c

^a*University Hospitals of Geneva, University of Geneva, Switzerland*

^b*Agfa HealthCare, Belgium*

^c*empirica GmbH, Germany for the DebugIT consortium*

Abstract. The concepts and architecture underlying a large-scale integrating project funded within the 7th EU Framework Programme (FP7) are discussed. The main objective of the project is to build a tool that will have a significant impact for the monitoring and the control of infectious diseases and antimicrobial resistances in Europe; This will be realized by building a technical and semantic infrastructure able to share heterogeneous clinical data sets from different hospitals in different countries, with different languages and legislations; to analyze large amounts of this clinical data with advanced multimedia data mining and finally apply the obtained knowledge for clinical decisions and outcome monitoring. There are numerous challenges in this project at all levels, technical, semantical, legal and ethical that will have to be adressed.

Keywords. Infectious disease, patient safety, semantic inter-operability, multimedia data mining, decision support, clinical outcome monitoring

Introduction

Building a safer and more efficient care system has become the most shared goal of all actors involved in healthcare. From a historical perspective, there has been an impressive shift towards awareness of the impact of errors in medicine in the last 25 years. In the early nineties, research papers and reports about patient safety, incident reporting and initial order-entry systems were published, mostly originating from academic settings. At about the same time, the first reports of the US Institute of Medicine (IOM) on computerized patient record systems stressed the ability of ICT-based solutions to improve the quality of care [1]. Ten years later, by the end of the nineties, a famous report of the IOM called attention to the wide prevalence of errors in healthcare [2]. While medical errors are under the spotlight, (re-)emerging infectious

¹ Corresponding Author: Service of Medical Informatics, University Hospitals of Geneva; E-mail: christian.lovis@hcuge.ch

diseases are becoming major challenges. Among them, the rapid development of anti-microbial resistances [3], the spread of nosocomial and other infections [4], the inadequate care and missing appropriate tools to lead the care system facing these new emergent problems [5] are major concerns. The issues around infectious diseases are strongly interrelated and have immediate and important effects on safety, quality of care and efficiency. In half a century of antibiotic use, new challenges have emerged: fast emergence of resistances among pathogens, misuse and overuse of antibiotics. Antimicrobial resistance results in escalating healthcare costs, increased morbidity and mortality and the (re-)emergence of potentially untreatable pathogens.

1. The project

Dedicated to infectious diseases, the DebugIT (Detecting and Eliminating Bacteria Using Information Technology) project aims at (1) detecting patient safety related patterns and trends, (2) acquiring new knowledge and (3) using this for better quality healthcare. A consortium of eleven partners has been built in order to gather scientific competencies in all domains involved, as well as to assure access to specific information of more than 2 millions clinical records.

Table 1. Consortium

Belgium	Agfa HealthCare
Bulgaria	GAMA/SOFIA Ltd.
Czech Republic	IZIP - Internetový Přístup Ke Zdravotním Informacím Pacienta
France	Institut National de la Santé et de la Recherche Médicale (with European Hospital George Pompidou)
Germany	empirica GmbH
Germany	University Medical Center Freiburg
Greece	Technological Educational Institute of Lamia
Sweden	Linköping University
Switzerland	University Hospital of Geneva
Switzerland	University of Geneva, Computer Science Department
United Kingdom	University College London

The project has a strong clinical lead guaranteed by a Clinical Advisory Board and a Scientific Advisory Board with European and American experts of the infectiology field and the scientific domains involved.

Outcomes and benefits, in clinical and socio-economic terms, will be measured. Results will be integrated into Clinical Information Systems (CIS) of participating European hospitals, industry and their clients, and become available globally through a European or global Disease Control Centre/Public Authority, also as Open Source solution. Advanced ICT applications and innovations concern the virtualization of the Clinical Data Repository through ontology and terminology binding and mediation, advanced data mining techniques, the use of machine reasoning related to real, point-

of-care patient data, as well as consolidation of all these techniques in a comprehensive but open framework. Output will be applicable to other clinical fields.

The concept developed as foundation for the DebugIT project addresses all of these issues in an operational manner with the ultimate goal to develop a new, highly advanced and pre-eminent tool aiming at producing a new and efficient weapon for the war against infectious pathogens across all health system actors and levels.

The overall project outcome will not only be a theoretical work and proof of concept, but also a practical implementation of a highly improved and advanced computerised system in the field of infectious disease treatment and antibiotics usage. This application, which, due to its generic conceptual base, should be easily expandable and adaptable to other similar medical application fields, will initially be evaluated by participating project partners, but should be made publicly available to other healthcare organizations soon after.

2. The Conceptual Framework

The conceptual framework of this project is an ever continuing iterative cycle, implementing the principle of translational medicine and true Evidence Based Medicine (EBM). Translational medicine makes the connection between medical research and clinical care by providing to research clinical data and providing the results of the research – the medical knowledge – as input for clinical care. While medical research is often focusing on prospective and tightly controlled studies, retrospective studies with access to huge amounts of data, just waiting to be analyzed, are a welcome addition to clinical research.

The framework can be broken down into several distinct steps (figure 1):

- 1- **Collect Data.** Clinical data will be collected and aggregated across different hospitals, countries, languages, information models and legislations, via advanced and commonly agreed data models (minimal data sets), standards and mapping algorithms.
- 2- **Learn:** Advanced data mining techniques on multimodal, multi-source, structured and unstructured data will detect patterns, relevant for patient safety and the treatment of infectious diseases such as: resistance of bacterias, adverse events and operational practices. This will result in new knowledge and new evidences for existing knowledge.
- 3- **Store Knowledge:** This knowledge will be stored, visualized, validated and aggregated together with pre-existing medical and biological knowledge (guidelines, regulations) to achieve a consolidated view on the needed knowledge, to be applied in the next step
- 4- **Apply:** Software tools will be integrated into the available clinical and public health information systems. Decision support tools will apply the generated knowledge and help the clinician to provide clinical care (choice, dose and administration of antibiotics for example). The knowledge will also be used to monitor the ongoing care activities and even predict future outcome to give additional feedback, both on individual patient and cohort level. This will allow healthcare providers and decision makers to take appropriate actions at various level of the healthcare system, including point-of-care, management or policy, and subsequently influence the future development of our health systems. Integration in existing CIS will enable

to record activities and results and thus make sure the necessary data are generated for a next cycle.

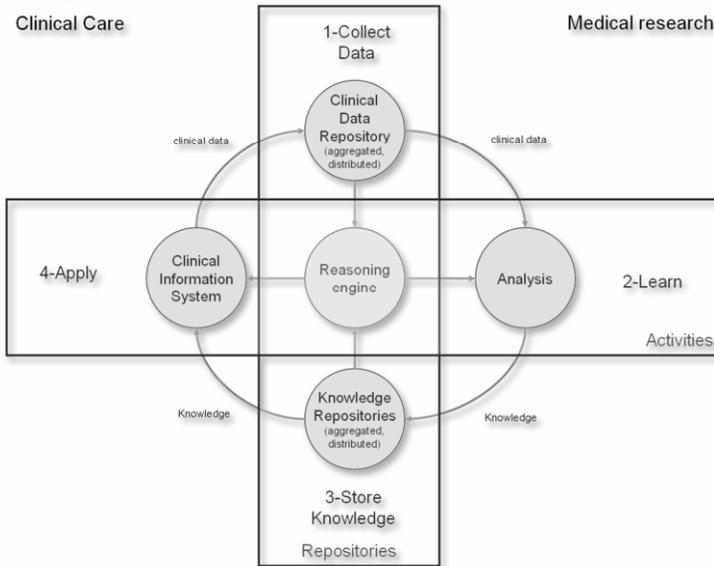


Figure 1. The Conceptual Framework

3. Technology and Architecture

To achieve his goals, the DebugIT project will make extensive use of clinical and operational information originating from running Clinical Information Systems (CIS) across the EU. The processes and entities are described in figure 2. Clinical data are accessible through a virtualized, fully integrated Clinical Data Repository (CDR).

The CDR will feature transparent access to the original CIS and provide data aggregations in local stores. The CDR is specifically tailored for knowledge discovery, featuring ethically sound, transparent access to data at or from the original CIS and/or collection and aggregation of data in a local data store.

Multimodal Data Mining (MDM) will have a strong focus on new fields of research doing mining on distributed storage, using highly advanced new text, image and structured data mining on individual patients as well as on populations.

New knowledge will be fed into a Medical Knowledge Repository (MKR) and mixed with domain knowledge coming from external sources (guidelines and scientific evidence). Innovative and user friendly knowledge representation paradigms will be developed in order to enable not only knowledge engineers but also clinicians to use the repository

After validating, this knowledge will be used by a decision support module (DSM) and monitoring tool in the clinical environment to prevent patient safety issues and report on them, both at the population and at the patient level for direct care.

Co-ordination and steering of both the analysis and the care process will be done by a performant and versatile reasoning engine.

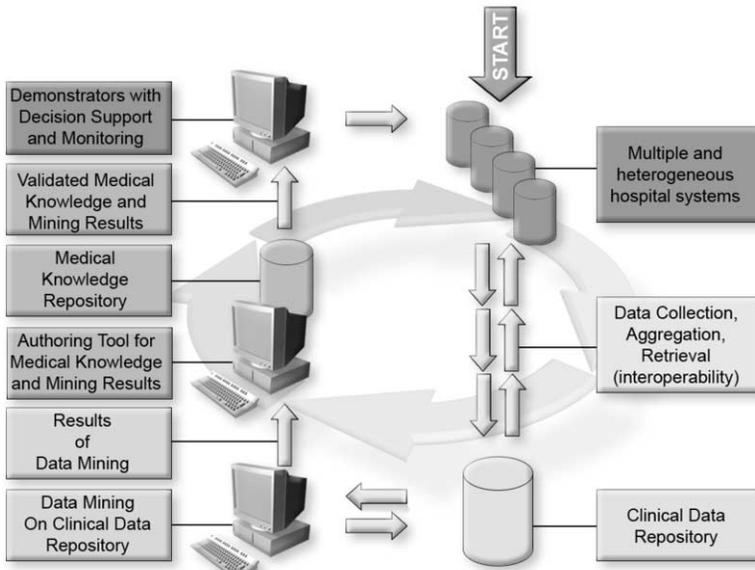


Figure 2: Overview of the project

4. Privacy

Strong attention is given to privacy concerns, taking into account the various legal and ethical frameworks to be met. Therefore, privacy is made a central part of the project *by design*, using a virtualized data repository without dealing directly with the original data. Identification elements provided by the clinical data repositories can be carried all along the process, blindly, in order to allow the original clinical information system to feed back decision support without need for patient identification.

5. Conclusion

There are huge challenges in the concepts and architecture, needing new and original solutions. Building a shared and distributed data repository with a common semantic infrastructure, dealing with privacy, etc, which will have to be addressed? The infectious model disease has been chosen as it allows addressing these issues with a pretty well defined domain of medicine. Infectious diseases have commonly fast course, identified causes and symptoms and well-defined treatments.

The DebugIT project is focused on using large existing and heterogeneous clinical datasets covering hundreds of thousands of patients from several clinical information systems in different European countries. DebugIt proposes to build an interoperability platform to populate a pertinent dataset about the infectious domain to achieve a very large common shared virtualized clinical repository that enables knowledge-driven data mining. This “semantic mining” will be based on innovative methodologies to deal with the characteristics of real world clinical data. A knowledge repository will drive

the data mining and serve as storehouse for the results. Finally, a decision support engine will exploit the aggregated knowledge to loop-back to the real world.

To achieve this system, several aspects will have to reach the frontiers of current state-of-the-art and beyond. Two strategies can be chosen for that. The first one is to invent something radically new. The second one consists of using all existing knowledge and methods, putting them together, and trying to build upon this base. For most of its research, the second strategy is the one chosen in this project, because operational results for clinical information systems must be available and sustain the DebugIT outcomes after the end of the project.

In order to meet these requirements, the project has been organized according to architectural component-based considerations:

- Interoperability Platform (IOP);
- Clinical Data Repository (CDR);
- Multimodal Data Mining (MDM);
- Medical Knowledge Repository and associated Knowledge Authoring Tool (MDR)
- Decision Support and Monitoring engine (DSM);
- Clinical applications.

This scientific and technical framework, associated with access to large amounts of clinical databases and led by experts in the medical field will lead to a serious advance in building a large IT infrastructure aiming at creating new knowledge in the field of monitoring, surveillance and efficient measures to fight infectious diseases.

6. References

- [1] IOM. The computer-based Patient record: An essential technology for health care. Institute of Medicine report. 1991, revised 1997.
- [2] Medicine Io, editor. To Err is Human. Building a safer Health System; 1999. National Academy Press.
- [3] Frazee BW. Update on emerging infections: news from the Centers for Disease Control and Prevention. Severe methicillin-resistant *Staphylococcus aureus* community-acquired pneumonia associated with influenza--Louisiana and Georgia, December 2006-January 2007. *Ann Emerg Med.* 2007 Nov;50(5):612-6.
- [4] O'Brien J A, Lahue BJ, Caro JJ, Davidson DM. The Emerging Infectious Challenge of *Clostridium difficile*-Associated Disease in Massachusetts Hospitals: Clinical and Economic Consequences. *Infect Control Hosp Epidemiol.* 2007 Nov;28(11):1219-27.
- [5] McAlonan GM, Lee AM, Cheung V, Cheung C, Tsang KW, Sham PC, et al. Immediate and sustained psychological impact of an emerging infectious disease outbreak on health care workers. *Canadian journal of psychiatry.* 2007 Apr;52(4):241-7.

Study on urban healthcare consumption in northern France

Julia SALLERON^a, Alain DUHAMEL^a, Charles SULMAN^b, Valérie DEKEN^a, Régis BEUSCART^{a,c}

^a*CERIM, University of Lille 2, Lille, France*

^b*City of Lille Councillor with responsibility for Health*

^c*Lille Regional University Hospital, Lille, France*

Abstract. In Lille (a large city in northern France), the municipal council has set up an Health Observatory in order to provide health information on the city's population, promote educational health policies, foster the development of preventive actions and improve access to the healthcare system. The Observatory works with agencies involved in health in France, such as the CPAM ("Caisse Primaire d'Assurance Maladie", the state health insurer). The purpose of the present study was to describe care consumption by the inhabitants of Lille's 12 districts on the basis of data supplied by the local branch of the CPAM. By using principal component analysis and a hierarchical classification tool, we established a typology of districts according to care consumption. The results of this study can be used to improve decision-making, elaborate better health policies and promote social actions.

Keywords: evaluation, decision support, classification, healthcare system, insurance, reimbursement

Introduction

Studies have shown that for some citizens, financial difficulties and casualization are obstacles to care access and continuity [1]. There are many disparities worldwide in terms of the consumption of medicines, and this is also the case in France. For example, narcotics consumption is twice the average of other European countries. In the Nord-Pas de Calais region (northern France), the rate of use of medicines for alcohol addiction is considerably higher than in the other French regions [2,3]. It is clear that care consumption and use of care services is not homogeneous. In Lille (a large city in the Nord-Pas de Calais region), an Health Observatory has been installed and works with local and national health agencies with a view to providing the health information that decision-makers and practitioners need in order to improve health policies. To date, studies performed on these topics have been confined to specific issues [4, 5]. This paper seeks to describe care consumption in various districts with different sociological profiles and establishes a typology of districts on that basis by using multivariate

statistical analyses (principal component analysis [6] and cluster analysis [7]). This method can play a useful role in elaborating health policies and social actions.

1. Method

1.1 Data collection

Data were supplied by the CPAM of Lille (the state health insurer). The study concerns persons affiliated with the general health fund and who were resident in the locality of Lille in 2004 and 2005. These inhabitants were spread across twelve districts. Originally, 154647 beneficiaries who had not changed address between January 2004 and June 2005 were included in the study. After an initial descriptive analysis, 114 beneficiaries were excluded from the study because they presented aberrant data. The study thus concern a total of 154533 beneficiaries, for whom the following information was supplied : sociodemographic data (age, gender, status in terms of special access to care and the latter's type), the number of consultations and visits to the doctor (both general practitioners (GP) and specialists), the amount spent on dental treatment (protheses, orthodontics) and the number of reimbursed boxes and prescriptions for all delivered medicines. Moreover, four specific medicines were identified according to the following therapeutic classes [8] : antibiotics, narcotics, tranquilizers and medicines for alcohol addiction.

1.2 Generation of indicators

Data were first aggregated by district and the following indicators were computed:

- the percentage of persons within a district having accessed a given type of care (dental treatment, protheses, specialists, etc.), the percentage of persons having received at least one reimbursement of any type of medicine and the percentage of persons having received at least one reimbursement of the four specific drug classes indicated above.
- the mean level of care consumption for the inhabitants of each district (20 indicators): number of consultations with a GP, specialist or dentist ; amount spent on protheses, orthodontics or dentist's fees and the number of reimbursed boxes and prescriptions of medicines received by the persons concerned.
- the mean age of the inhabitants in each district, as well as the percentages receiving a guaranteed minimum welfare allowance ("RMI"), free state health insurance ("CMU"), state medical aid for non-French citizens ("AME") or a chronic disease allowance ("ALD").

In all, we used 39 indicators to generate a summary of care consumption and the use of care services for each district. We then perform principal component analysis (PCA) in order to select the most relevant indicators. This analysis enabled us to select 20 indicators for the remainder of the study.

1.3 Typology of districts using cluster analysis

A hierarchical cluster analysis (HCA) was performed to group together districts in which the inhabitants displayed similar behaviour in terms of care and medicines consumption. The HCA is a multivariable statistical method based on the calculation of distances between the districts; the goal is to identify clusters of districts having closely related characteristics for the 20 analyzed parameters. The number of clusters was determined by using a consensus between three statistical parameters: the cubic clustering criterion (CCC), the pseudo F and the pseudo t^2 (the SAS *cluster* procedure). In addition, we performed a factorial discriminant analysis (FDA) in order to check the quality of cluster separation. The mean values of the indicators were calculated and used to compare the clusters in an analysis of variance (ANOVA). These two analyses (FDA and ANOVA) enabled us to identify the indicators which best separated the clusters.

1.4 Radar charts

We used radar charts to represent the mean differences between the districts and the clusters, according to the indicators. Each indicator was transformed into a z-score, which is the current indicator value minus the overall mean divided by the standard deviation. The mean values for the overall population are thus always equal to 0.

2. Results

Given that use of care services differs strongly between children and adults, the results presented below concern the 125265 inhabitants over the age of 15. The mean age of this population was 42. In this population, 3% received the guaranteed minimum welfare allowance ("RMI"), 9% the chronic disease allowance ("ALD") and 12% the free state health insurance ("CMU").

2.1 District-by-district analysis

We analyzed twelve districts in the city of Lille. Figure 1 presents the results obtained for one district, called District 1, a residential area with approximately 20000 inhabitants. This district is characterized by very high unemployment rate (over 20%) and a high proportion of low social classes (35%). Many of the area's inhabitants are first-, second- or third-generation immigrants (over 25%). Radar charts enabled us to characterize this district in terms of care consumption and the use of care services.

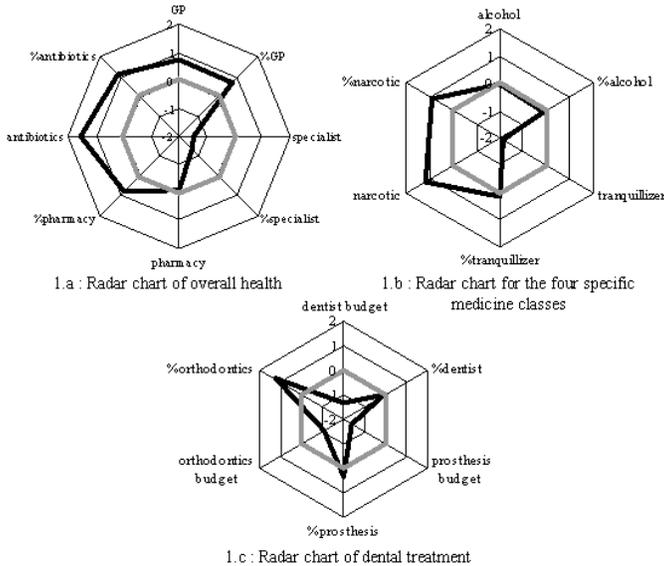


Figure 1 : Radar charts for the District 1

The percentage of the district's inhabitants receiving medicines (85%) was higher than that of the general population (81%) but the mean number of prescriptions delivered was very similar (18.2 versus 18.3 in the general population) (*Figure 1.a*). The percentage of persons having consulted a GP was also higher (82.6% versus 78%), as was the mean number of consultations (12 versus 11). In contrast, the percentage of persons having consulted a specialist was much lower than in the general population (42% versus 51%) and the mean number of specialists consulted was lower than in the general population (1.7 versus 1.9). The percentage of the district's inhabitants receiving reimbursed antibiotics was higher than in the general population (59% versus 54%), as was the mean number of prescriptions delivered (3.2 versus 2.9).

Figure 1.b concerns reimbursed prescriptions of narcotics, tranquillizers and medicines for alcohol addiction in District 1. In this district, the percentage of persons taking treatments for alcohol dependence was the same as in the general population (0.9% versus 1%) and these individuals did not consume more than the inhabitants of the other districts (4.23 delivered prescriptions in District 1 versus 4.26 in the general population). District 1 featured a higher proportion of persons taking narcotics than in the other districts (1.54% versus 1.33%) and these individuals received more prescriptions on average than those in the other districts (16 versus 13.7). In terms of tranquillizers, District 1 has the same percentage of consumers as in the general population (29% versus 28.7%) but these individuals receive fewer prescriptions on average than in the other districts (10.3 versus 12.2).

The mean amount spent on dental treatment by District 1 inhabitants was well below the average for the general population (€374 versus €421 for dental treatment, €774 versus €843 for the orthodontics and €668 versus €836 for prostheses) (*Figure 1.c*). Nevertheless, the level of use of this type of care service was the same as in the other districts and was even higher for orthodontics (1.9% versus 1.4%).

2.2 Typology of districts using the hierarchical cluster analysis

Using the above-described method, the 12 districts of Lille were distributed into three clusters (Figure 2.a). ANOVA and FDA enabled us to determine the 11 indicators which were the most effective in separating the clusters. Figure 2.b shows the radar chart obtained for the three clusters. One can note that the reimbursed medicines and the use of care services differ greatly from one cluster to another.

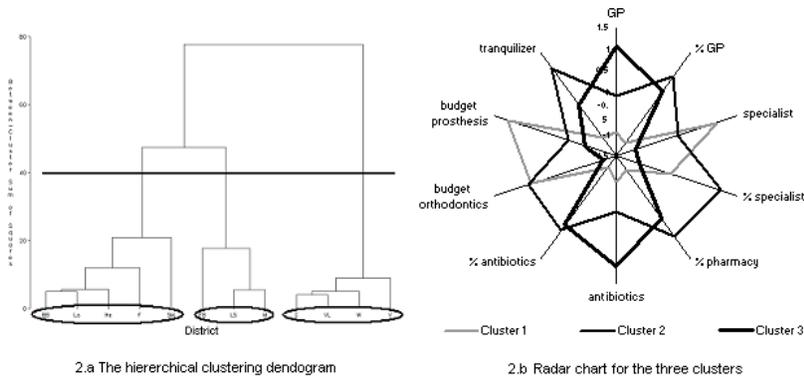


Figure 2 : Results

The above-described District 1 is part of cluster 3, formed by 3 districts which differ significantly from the general population. Firstly, the percentage of the inhabitants consulting a GP was higher than in the general population (81% versus 77%), as was the mean number of doctors consulted (13 versus 11). In contrast, few persons consulted a specialist (44% versus 51% in the general population) and the mean number of specialists consulted by these individuals was lower than that in the general population (1.5 versus 2.0). The mean amount spent on dental treatment was lower than for the overall population (€753 versus €836 in the general population for prostheses and €776 versus €841 for orthodontics). Lastly, the percentage of persons having been reimbursed for prescriptions of antibiotics (56.4%) and the mean per capita number of prescriptions delivered for these individuals (3.1) were higher than in the general population (with 2.6 delivered prescriptions and 52.8% by person respectively). The percentage of people having been reimbursed for prescriptions of medicines for alcohol addiction was higher than in the general population (1.4% vs 1%).

3. Discussion – conclusion

Although (in France, at least) municipal councils do not have direct administrative responsibility for the provision of healthcare services, an increasing number are becoming more and more concerned in the field of public health. Here, we have presented a method for characterizing use of care services and healthcare professionals

for the inhabitants of 12 districts in the locality of Lille. This method also allowed us to group together districts having similar characteristics in terms of use of care services and consumption of medicines.

This decision-making tool enables us – in the short-term – to draw up an objective overview of the nature and the extent of disparities in term of care consumption. Concretely, it informs all the decision-makers of the needs in health care. The impact of this kind of information is not immediate because it needs a reflection of the local actors to ensure its interest and its relevance. In the mid-term, it will be a question of developing a local plan in public health and of planning programs of intervention. For example, our study reveals that in some districts, people use less dental treatments. To level this disparity, an information and detection campaign could be done in the schools of this district to detect a lack of hygiene or dental problems.

This kind of database contains information on reimbursement of all care (other than during hospitalization) for all individuals, making it possible to obtain relevant information about care consumption. However, the study is limited by the nature of the data, especially for medicines : the fact that the cost of a prescription has been reimbursed to an individual does not necessarily mean that the individual has actually taken the medicine. Furthermore, (non-reimbursed) self-medication is not taken into account.

In terms of the causes of the observed inter-district differences, this study raises more questions than it resolves. A factor which is often put forward is casualization. People with the lowest income are generally more likely to suffer from certain diseases (for example, 11% of the poorest people suffer from tooth decay, versus 6% for the rest of the population [9]). Low-income individuals are also less likely to visit the doctor in general and specialists in particular. The method described in this paper could be a way of answering this question : the same type of study could be performed by replacing the districts by the level of casualization.

References

- [1] 2006 report from the Healthcare Access Observatory of the French branch of 'Médecins du monde', October 2007.
- [2] Lecadet J, Vidal P, Baris B, Vallier N, Fender P, Allemand H and Medipath, Psychotropic Medications: Prescriptions and Use in Metropolitan France, Medical magazine of Health Insurance, 2003, 34 (4) : 233-248
- [3] Mancini J, Thirion X, Masut A, Saillard C, Pradel V, Romain F, Pastor MJ, Coudert C, Micallef J, Anxiolytics, hypnotics, and antidepressants dispensed to adolescents in a French region in 2002, Pharmacoeconom Drug Saf, 2006, 15(7):494-503.
- [4] Vignau J, Duhamel A, Catteau J, Parquet PJ., Practice-Based Buprenorphine maintenance treatment (BMT): how do French health care providers manage the opiate-addicted patients ?, Journal of Substance Abuse Treatment, 2001, 21(3): 135-144
- [5] Fuhrman C, Sarter H, Thibaudon M, Delmas MC, Zeghnoun A, Lecadet J, Caillaud D, Short-term effect of pollen exposure on antiallergic drug consumption, Ann Allergy Asthma Immunol, 2007, 99(3):225-31.
- [6] Morrison DF, Multivariate Statistical Methods, Second Edition, New York: McGraw-Hill Book Co, 1976
- [7] Hartigan JA, Clustering Algorithms, New York: John Wiley & Sons, 1975.
- [8] European Pharmaceutical Market Research Association (EphMRA)
- [9] Saint Pol T, The health of the poorest, INSEE Première No 1161, 2007

Cross-Border collaboration between Greece and FYROM: Mobile Healthcare Provision

Stergiani SPYROU^a, Dimitrios VARTZOPOULOS^a, Panagiotis BAMIDIS^b
Nicos MAGLAVERAS^b

^a*3rd Regional Healthcare Authority (Macedonia), Greece*

^b*Lab of Medical Informatics, The Medical School, University of Thessaloniki, Central Macedonia, Greece*

Abstract. Introduction of eHealth tools and applications denotes the new era in health care sector and especially in health care networks. The telemedicine applications in cross-border areas, referred as a Cross-Border Health network, serve the improvement of the quality of life for the population in cross-border areas. In this work a framework for such a network concerning the collaboration between Greece and FYROM is described. The network is in the first phase of design and is expected to be implemented within the next year. The requirements, the restrictions and the design of the network has been defined by the healthcare professionals and it staff that participate in the project. The results, so far, reveal the acceptance of the system from the staff of the healthcare organizations, while detailed results for the performance of the system will be available in the first quarter of the next year. The work denotes the successful efforts for the development of Cross-border Health Networks.

Keywords. eHealth, Cross-Border Health Network, Telemedicine

Introduction

eHealth tools and applications that improve the performance of health care organizations are the chronic disease management, eBooking, eLearning, ePrescribing, Electronic Health Record, Hospital Information System and other. eHealth applications that improve the services of several cooperating healthcare organizations are Cross Border Health Networks, Regional Network, National Network, Telecare and Telemedicine [1]. Within the scope of the cross border collaboration that serve the INTERREG III program is included the effort of the 3rd Regional Health Authority (RHA¹) of Greece to apply eHealth applications in order to serve the collaboration of Greece and the Former Yugoslav Republic of Macedonia (FYROM) in the axe of improving the quality of life and health care provision.

¹ Abbreviations used in this paper: RHA stands for 3rd Regional Health Care System Authority of Greece, COPD stands for Chronic Obstructive Pulmonary Disease.

3rd Regional Health Authority manages 17 hospitals, 17 Health Units and 166 district health units. Penetration of Information and Communication Technology (ICT) infrastructure in hospital is extensive as there is Management Information System (MIS) in almost all hospitals. The Greek public network for providing internet access and telephony service has recently installed in most of the health units. Most importantly the 3rd RHA runs a project for the “Integrated Regional Health Information Network” [2] to implement information system supporting administrative and clinical services in hospitals, health units and Regional Health Authority.

On the other hand, INTERREG III is a Community initiative which aims to stimulate interregional cooperation in the EU [3]. It is financed under the European Regional Development Fund (ERDF). The initiative is designed to strengthen economic and social cohesion in the EU and special emphasis has been placed on integrating remote regions and those which share external borders with the candidate countries. Greece and FYROM are such two countries that share external borders.

eHealth projects introduce innovative technologies that enhance the cross-border collaboration of the two countries and enhance the level of health services for the population of border districts. With the aim to develop infrastructure for medical diagnosis and treatment along with collaboration of medical professionals of the two countries the framework of the project has been designed. The framework could be applied to ehealth projects that support collaboration of regions and lead to the development of Cross-border Health Networks. The proposed network concerns the collaboration of health institutions of Greece and FYROM, for the confrontation of asthma and Chronic Obstructive Pulmonary Disease (COPD) and other chronic respiratory illnesses. The telemedicine services concerns medical diagnosis at distance, education and prevention of illnesses. Moreover, the creation of a web portal for the education of patients on issues regarding asthma is included. The work will also focus in the preventive medicine to people with respiratory diseases that are found in district regions. The project is in the first phase – including the requirement analysis – and by the end of the first quarter of the next year detailed results of the project will be available.

1. Background

Primary health care provision is of paramount importance in Regional Health Care Systems. GPs and other doctors support the health care provision. However, health units do not provide medical services regarding specific medical specialities such as pneumonologists or other. E-health provides the technical means for the GPs to be in conduct with doctors with medical specialities in hospitals either with tele-monitoring or tele-consulting and other applications of eHealth. The ICT infrastructure gives the means to the patient or to the doctor to collaborate with the expert, avoiding the patient flow from rural areas to the urban.

In 2004 the 3rd RHA has engaged a survey regarding the attitude of staff of four health units towards the use of ICT and the results were positive. The received questionnaires were 90 and the staff that participated in the study included doctors, nurses, administrative personnel and other. 41.83 % of the staff had already used Electronic Medical Record (EMR) software, 66.67% accomplished part of their work using Internet, 27.78% used Internet for commercial transactions like banking transactions. As a conclusion most of the staff regarded that the use of ICT

infrastructure would make their work easier, would improve the performance of the provided services and generally would provide a reliable way of working. Based on the positive attitude of personnel along with the benefits of the usage of eHealth tools to improve the health care services, the 3rd RHA has accomplished successfully some telemedicine projects and is developing some others. In the next paragraphs, short descriptions of two of those are given.

A telemedicine project included the integration of GPs in five rural Health Units with General Hospital and Expert Doctors for the need of following up the chronically ill citizens located in remote regions in Region of Central Macedonia, Greece had implemented successfully. The implementation has been based on the Mobinet service [4]. Communication of health centers and hospital was achieved by micro-telemedicine devices, it-telecom components (mobile phone etc), wireless communication networks and cellular telephony network. The Clinical Records (EMRs) were kept in a web server and access provided to physicians, patients and hospital staff. The EMR is enriched with vital signs like ECG, sets of pulmonary function parameters, heart rate, diastolic & systolic blood pressure, SPO2 and blood.

The results were impressive concerning the adaptation of doctors and patients to the usage of the system. Although none of the doctors had previous experience using telemedicine systems, they were all well-educated both in the ICT infrastructure and the usage of mobile medical devices. They all used the system for approximately 6 months and 981 sessions were accomplished. Each session included the patient data entry to the mobile device and the transmission of data via GPRS to the expert doctor. Expert doctors received the data from each session and send the clinical note either via the web-based application and internet or sending sms message to the GP. The system was used by the doctors several times per week as it was used for the follow-up of chronic disease patients or for children with the suspicion of asthma. GPs spend approximately 6.5 minutes per session for the usage of the system. However the expert doctors spent much time on the results of medical act in order to give his opinion.

Another project that is under development is the tele-psychiatry that is within the framework of project of "Integrated Regional Health Network" of the 3rd RHA [2]. In this project a tele-consultation scenario for psychiatric patients has been designed and developed. The scope of the project is to provide means for the follow-up of psychiatric patients from the remote health centers during their rehabilitation along with tele-consultation services for patients visiting the health centers. Within the range of the project is the communication of medical personnel of two hospitals in the region of RHA, six district health centers and units and two mobile units. The project has not yet finished and results are expected to be announced within the next year.

Several other telemedicine projects has been designed or implemented by the health organizations of RHA. However, a cross-border eHealth project has not implemented so far.

2. Requirements and design of the system

The healthcare organizations that participate in the Cross-border Health Network system include the two hospitals of Thessaloniki, namely the General Hospital of G.Papanikolaou and General Hospital of Papageorgiou and more specifically their pulmonary clinic. In one of them the telemetry central station will be settled and will be in coordination with the system located in the 3rd RHA head office. Five health units

and district units from the region of Central Macedonia will also participate in the system as the remote units for the project. On the side of FYROM a hospital with three local points will also act as remote units for the project. The requirements are expressed with the scenario that is going to be implemented in the project and is presented in this section.

The proposed services concern the education of patient, via specialised web portal, where educational material will included for the confrontation of asthma and chronic obstructive pulmonary disease (COPD). The content of the portal will be available in three languages and it will be also used by the population of the regions that participate in the work. Also, the services concern the follow-up and convenient confrontation of any crises of chronic patients with respiratory diseases.

A central station of the telemetry service will provide the services to all remote units of the cross-border network. Hospitals and health units will be equipped with suitable devices (pcs, spirometers etc) for measurement of vital signals of patients. In the remote units the medical personnel will record the measurement of violent expiry (FVC) of patients with suitable appliances and then the measurements will be dispatched via telecommunication network (telephone, GPRS, etc) in the telemetry central station. In the case of chronic illness the health care professional in collaboration with the patient will establish a plan of follow-up of his health. At the end of installation of equipment and the test operation of services, the education of personnel to the use of equipment will follow. At the same time, the system will be installed in neighbouring municipality of FYROM, aiming to exchange of know-how and scientific knowledge for the joint confrontation of asthma and COPD. The hospital of FYROM will provide local points to use the service of telemetry. The healthcare organizations of Greece and FYROM will present the results of the use of the telemetry system in the regions and will study and make a comparative analysis of results.

The restrictions of the system supporting the cross-border network are: (1) lack of specialized personnel to the usage of telemedicine systems, (2) lack of personnel to support the central station at a 24hour base, (3) bureaucracy in the procedures to deploy extra personnel and (4) integration of EMR with other eHealth systems [4] that are installed in health units or hospitals. Under those restrictions that apply to many telemedicine or generally eHealth projects, the framework of the system was design as is briefly described in the next section.

3. Framework

The system that is designed is shown in figure 1. In figure 1 the personnel is shown along with the equipment that the personnel is going to use. The personnel could be public officers or external contractors. For the needs of the project the external contractors of the system will be the following: (1) consultant of technical support: for administrative, educational and other issues (Actors1), (2) health care professionals to support the telemetry central station and the doctors in primary health units (Actors2), (3) health care professionals and it staff to coordinate and support the creation of educational material for the web (Actors3), (4) administrative personnel to support the processes (Actors4) and (5) IT companies to provide and support the project (Actors5)

We are going to refer to each of the above groups with the description shown in parenthesis. For example Actors4 is the shortcut name for the group of administrative personnel to support the processes of the system. Actors2 and Actors4 can be public

officers and not external contractors. The phases of the project are five: (1) PHASE A: Requirement analysis and design of applications, (2) PHASE B: Development, parameterisations of applications, tests and integration of applications, (3) PHASE C: Supply and installation of equipment (pcs, microdevices etc), (4) PHASE D: Period of Pilot Operation - Education of users, (5) PHASE E: Period of SLA (Service Level Agreement). Several deliverables by the external contractors and public officers are included in their work. The most important include the actions in Table 1:

Table 1: The actions and actors in the system

AXIS	Actions in the axis	Involved Actors
Organizational matters	Reports including:	
	Legislation for security and other matters	Actors1
	Organizational matters	Actors 1, Actors4
	Functional requirements	Actors3, Actors5
Architecture of the cross-border network	Educational needs	Actors2, Actors3
	Design of architectural model of telemedicine services	Actors1,
	Technical characteristics of infrastructure	Actors2,
	Design of educational material	Actors5
Financial Planning	Cost-benefit analysis report	Actors2, Actors3
Dissemination	Reports including:	Actors4
	Acceptance of the system by users, patients, doctors	Actors1, Actors3
	Reliability model of the telemedicine system	

The citizens that can profit from the proposed service of telemedicine are those with chronic respiratory diseases, as well as those with particular needs of follow-up of of their health as: a) Children with asthma b) Old men that should watch the development of their health, g) Persons with special needs, e) smokers d) persons with heart failures. The scientists of two regions might examine joint incidents that need investigation, while at the same time is created a medical database for the populations of cross-border regions, allowing thus the conduct of comparative studies and statistical analyses.

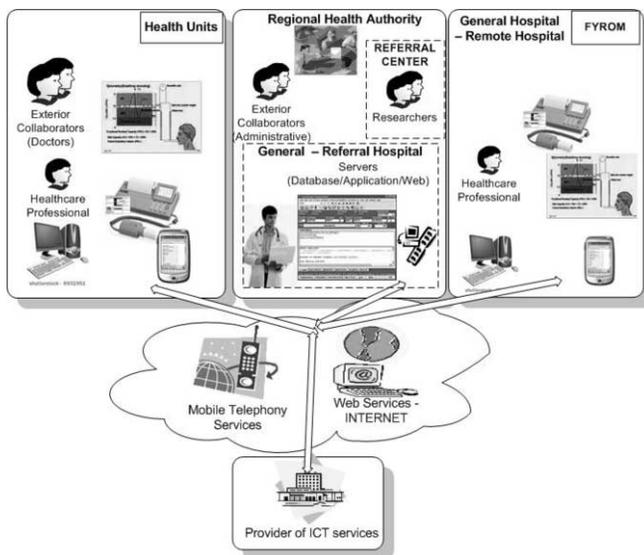


Figure 1: The personnel and equipment of system

4. Discussion and Future Work

In the cross-border area, efforts are made to cooperate in matters of improvement of quality of life. Increased cooperation in the sector of new technologies (especially as far as health is concerned) will lead to the establishment of new ways to tackle local problems affecting the region's development.

The integration of telemedicine in the sector of public health through the proposed work promotes the level of provided services of health and contributes in the quality of life of population and exchange of knowledge for health care professionals. The proposed service of telemedicine ensures the provision of specialized medical services of patient near the place they live, avoiding frequent visits in the urban centre - hospital.

Cross-border network offer the opportunity to people using the tools of eHealth to be evaluated by specialised doctors. Patients avoid frequent visits in the urban centre - hospital. Patients also undertake a more active role in the follow-up of their health. The doctors having data from their patients frequently can offer better treatment. The hospital ensuring reliable control of patient from distance can decrease the hospitalisation per year, decreasing the incidents that need internal hospitalisation, decreasing the disposal of resources. Also, the hospital is able to offer medical services in wide teams of population, which could not previously approach because of geographic restrictions. Moreover, the forecasted cross-border action and diffusion of results of work in adjacent regions will contribute in the assimilation of technology and promotion of proceeding of telemedicine, demonstrating the effect of eHealth in primary care.

The expected results are summarised in the followings: overcome geographic restrictions for diagnosis by distance, improvement of quality of provided services of preventive diagnosis and medical care, reduction in the costs of resources, reduction of expenses of patients, reduction of time and costs of hospitalisation.

Acknowledgments

This work was supported and benefited from INTERREG IIIA Greece-FYROM (Decision 300658/15-03-2006; Interregional Health Network). Doctors and IT and Administrative personnel and are gratefully acknowledged for their ideas, their support and their work in the project.

References

- [1] eHealth tools & services, "Needs for the Member States", Report of the WHO Global Observatory for eHealth, World Health Organization, <http://www.who.int>, accessed 17-03-2007
- [2] Information Society SA, "Healthcare Information System for the 2nd Regional Healthcare Authority of Central Macedonia", *Request For Proposal co-funded by the 3rd CSF under the EU decision C(2001)551/14-3-2001*, Greece, 15 May 2003 [In Greek].
- [3] Community Initiative INTERREG III, <http://www.interreg.gr>, accessed 10-11-2007
- [4] S. Spyrou, A. Berler, P. Bamidis, "Information System Interoperability in a Regional healthcare System Infrastructure: a pilot study using healthcare Information standards", Proc. Of MIE 2003, Saint Malo, France, pp 364-369, IOS Press, 2003.
- [5] P. Angelidis et al., "MOBINET: e-Wellness services fostering citizen empowerment", proc. in The International Special Topic Conference on Information Technology in Biomedicine (ITAB) 26-28 October, 2006 (<http://www.medlab.cs.uoa.gr>)

9. Privacy and Security

This page intentionally left blank

Secure dissemination of electronic healthcare records in distributed wireless environments

Petros BELSIS*, Dimitris VASSIS, Christos SKOURLAS, Grammati PANTZIOU
Technological Education Institute of Athens

Abstract. A new networking paradigm has emerged with the appearance of wireless computing. Among else ad-hoc networks, mobile and ubiquitous environments can boost the performance of systems in which they get applied. Among else, medical environments are a convenient example of their applicability. With the utilisation of wireless infrastructures, medical data may be accessible to healthcare practitioners, enabling continuous access to medical data. Due to the critical nature of medical information, the design and implementation of these infrastructures demands special treatment in order to meet specific requirements; among else, special care should be taken in order to manage interoperability, security, and in order to deal with bandwidth and hardware resource constraints that characterize the wireless topology. In this paper we present an architecture that attempts to deal with these issues; moreover, in order to prove the validity of our approach we have also evaluated the performance of our platform through simulation in different operating scenarios.

Keywords. Telematics, Distributed systems, Networks, Security.

Introduction

Wireless technologies have gained wide acceptance while they have been integrated in many environments our everyday life. As mobile devices become more powerful, they are integrated as a non high-cost solution in many environments. Among other reasons for their wide acceptance we can distinguish their relative easy integration in many types of environments as well as the continuous raise in computational power of handheld devices. The medical domain can suffice a lot by their deployment, since the use of handheld devices may provide doctors with accurate information from every place within a wireless environment which usually covers the clinic in which they belong to. Therefore, in contrast to past practices, the doctors do not need to approach a steady point (for example their office) and get access to medical data (concerning their patients) using their pc and by accessing a specific application or by logging to a portal; instead while carrying their wireless device they can use it to acquire on site and in

* Corresponding author: Petros Belsis, Dept. of Informatics, Technological Educational Institute of Athens, Ag. Spyridonos Street, 12210, Aigaleo, Athens, Greece, Email: pbelsis@cs.teiath.gr

timely accurate manner access to medical information, realizing thus the anytime-anywhere access to medical information paradigm.

Among the main design and implementation challenges we can distinguish:

- The capability to provide information to doctors independently of their exact location;
- Achievement of information integration using interoperable standards for medical information storage and exchange;
- The ability to ensure that no sensitive medical information will be disclosed to unauthorized parties.

In this paper we present an architecture that allows authorized medical personnel to access medical information while moving within their clinic; therefore they can gain access to their patient's medical records using an architecture that utilizes standardized technologies for processing, storage and exchange of information. Lightweight yet robust encryption and authentication techniques ensure that no breach of confidentiality happens. For our specific design choices, we measure the information payload of HL7 messages and perform simulation tests that show the validity of our approach. We also describe our benchmark methodology through which we have determined the effect of performance issues such as determination of the maximum number of concurrently sent HL7 messages, considering at the same time the encryption overhead. The rest of the paper is organized as follows: Section 1 presents related work in context. Section 2 describes in brief the main design principles which rule our approach. Section 3 presents our architecture, and an overall usage example scenario. In Section 4 simulation results are outlined, by which our approach is evaluated and validated. Finally, Section 5 concludes the paper and provides the directions of our future work.

1. Related Work

Electronic health records management attracts increasing interest and sets the scenery for the establishment of ubiquitous acquisition of medical information, achieving thus the provision of improved health services. Lately several projects have focused on providing efficient solutions to various aspects of medical records management, leading also to the development of improved e-health services [7][9][10].

Wireless mediCenter [3] is a system for management of electronic medical records and delivery through secure LANs or high-speed wireless connections. It provides different portals for doctors and patients in order to achieve classification of access permissions. The restriction though to connect through the portal is a serious burden to the user.

The m-Care project [4] aims at providing secure access through a WAP based architecture. Users and access rights related information is kept in an MS-SQL Server database. In our approach we have enabled a policy based approach which facilitates interoperation with other systems while it also provides a highly distributed nature to our system. PatientService [8] is a trust-based security architecture that enables medical records management in pervasive environments. In this approach access to medical information is provided to a set of users which hold a PDA that keeps the policy in a smart card. In our approach we issue the request from the PDA while the policy evaluation is not performed by the PDA itself. Moreover we attempt to evaluate our approach by performing simulation experiments.

2. Design requirements for wireless medical infrastructures

Pervasive infrastructures are characterized by rapid changes in their topology. Devices on the other hand, are characterized by the low computing resources and power. Medical information on the other hand is highly sensitive; thus, we have to design our system so as to demand as less processing and network bandwidth resources, without though decreasing our strict privacy requirements. Among the main requirements for our architecture we can distinguish:

- **Privacy preservation.** Unauthorized disclosure of medical information may lead to disastrous results. EU and US legislation force towards privacy preservation of medical data. Except from protecting appropriately medical databases, transmission of medical information should also be performed in a reliable and secure manner. We have thus employed efficient encryption techniques based on both symmetric and public key cryptography methods, so as to achieve data protection without demanding excessive processing power. In order to transmit data wirelessly we first exchange a shared key using strong encryption based on signing the messages with the private keys of the two parties, and then continue using shared key encryption so as to achieve a lightweight implementation.
- **Network topology instability:** Pervasive infrastructures are characterized by node mobility as well as node failure; In order to enable constant connectivity for as long as possible, we have decentralized many of our processing and communication tasks avoiding thus the existence of points of failure. Towards this direction we have adopted the DLS (distributed lookup server approach) [5] according to which a number of nodes act collectively as a centralized node. When a node is about to stop transmitting, it passes all of its information to its neighbors.
- **Access control management:** In order to apply access control, we have adopted the Role Based Access Control Model (RBAC), due to its simplicity and wide acceptance as a security standard. Access Control is performed in the medical database using a policy based approach [6]. A policy approach allows determination of privileges according to business roles, accordingly these privileges may be encoded in a suitable policy language and each request is directed towards special purpose modules, which reason over a specific request and either authorize or reject the request.
- **Interoperability:** In order to enable interoperation of our system with other medical systems and architectures we have adopted the HL7 standard for information encoding and exchange. For secure transmission and in compliance with the guidelines of the HL7 that instruct the use of secure protocols, one of the IP Security protocol (IPSec), Secure File Transfer Protocol (SFTP) or Secure Socket Layer Protocol (SSL) can be used for encrypting medical records.

3. System Utility Scenarios

We can consider the following scenario: A doctor within the ward he/she belongs approaches a patient's bed and wants to access some basic information regarding the patient's medical history. The doctor sends a request from his/her PDA to retrieve the

data from the medical database. The medical database is kept in a device with more processing power which acts as a medical server. In order to avoid single point of failure we have adopted a distributed approach considering caches per sub-domain. Moreover, in order to lower the network resource consumption we have adopted the use of ontologies one per sub-domain (fig. 1). Thus, when a device from one domain wants to query for specific information it firsts examines the ontology of the domain under consideration. For example if we are interested in medical records concerning hematological data and the ontology of the domain does not specify such terms, then it is useless to query the medical servers for information of that kind regarding a specific patient. When a request is sent to the server, in order to authorize or not the request, the server needs to identify the doctor's identity as well as to evaluate the permissions which have been granted to the doctor. First it requests a validation of the doctor's id. This can be done using public key encryption techniques. The doctor's private key is stored in a smart card inserted in the PDA. Using the doctor's public key and the server's private key, the two parties may authenticate each other and they can exchange a (shared) session key which will be used to encrypt all further communications. This is being done since use of private key encryption techniques for the transmission of all messages would demand a lot more computational resources. The doctor's device also is able to identify whether it resides within the clinic or whether it resides in an unknown environment with the aid of a beacon which sends signed messages identifiable by the doctor's device when compared to a number of stored signed (within the smart card) messages. Thus, we prevent unauthorized transmission or reception from the device when it resides outside pre-settled space boundaries. After authentication has been performed and the session key has been exchanged, all communication can be encrypted end to end from the medical database to the doctor's device using SSL. When a new request is sent to the medical database, the policy module is invoked, which examines the request, the requester's role and the privileges which have been recorded in the policy. This procedure is supported by most of modern devices which handle effectively at least 128-bit encryption.

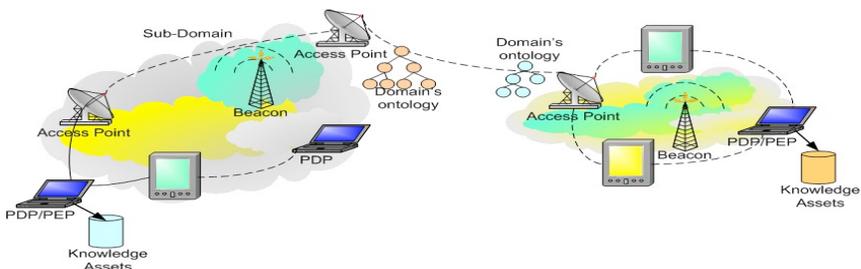


Figure 1. Pervasive medical domain overall architecture. The presence of a beacon on each domain can notify the device about the local policy enforced and direct easier user-authentication. Devices of higher processing power may work as a medical server. In the figure we see also PDA's of doctors participating in the wireless framework.

4. Performance Evaluation

In order to test the validity of our approach we have evaluated its performance through simulation in Pamvotis WLAN simulator [1]. We assume an IEEE 802.11 wireless channel of 1Mb/s, which is capable of covering a range of up to 300m of indoor

environments. The IEEE 802.11 protocol is suitable for HL7-based applications, as it supports high data rates and combines the advantages of mobility and packet switching, making it suitable for IP-based mobile devices such as PDAs and 3G/4G mobile phones.

Considering an average HL7 message, we assume that the application packet payload size obeys a uniform distribution with a mean value of 280Bytes. Concerning the packet overhead added from the SSL protocol, this is about 16% of the packet payload [2], meaning 48Bytes. Adding the TCP overhead (32Bytes with timestamps included) and the IP overhead (20Bytes), we have an IEEE 802.11 MAC datagram of 380Bytes. Finally, we assume that the number of messages each doctor sends obeys a Poisson distribution with a mean of one message every two minutes. The simulation results concerning the above configuration are outlined in the rest of this section. 2 depicts the aggregate traffic (system throughput) generated by HL7 messages for a network consisted of 600 users.

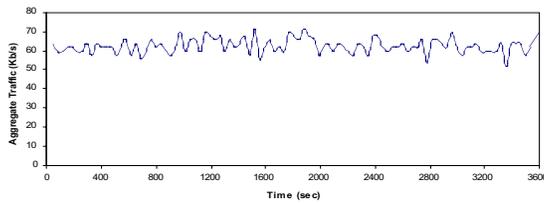


Figure 2. Aggregation traffic versus simulation time.

As we can see, the traffic is minimal compared to the wireless channel capacity. Hence, if other applications run on the same network (e.g. medical image downloading or interactive applications such as VoIP), our proposed architecture does not affect their performance.

Figure 3 depicts the service delay for a network consisted of 600 users. What we mean by service delay is the delay from the moment a user sends an HL7 message until the moment he receives another HL7 message from the database server, containing the information requested. Note that the processing delay (e.g. delay faced on the database server for performing the database transaction) is not taken into account.

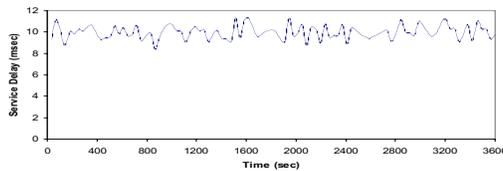


Figure 3. Service delay versus simulation time.

We observe that the service delay is less than a second, even for the best effort class considered in our case. Finally, Figure 4 depicts the service delay versus the HL7 message size in ASCII characters.

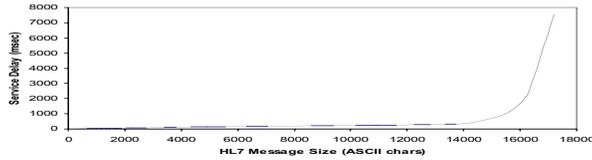


Figure 4. Service delay versus message size.

As we can see, the delay is low for messages up to 1500 characters, allowing a large amount of information to be contained in the messages. For messages up to 1500 characters the system gets saturated and the delay increases significantly. For example, the service delay for a message consisted of 17196 characters is about 7.5sec.

5. Conclusions

Pervasive environments offer new possibilities for efficient and better e-health services. Still, the nature of medical environments imposes many restrictions and demands consideration of several factors, such as: considering the limited computational, power and network bandwidth resources and of course security. In this paper we have presented an architecture that allows authorized medical personnel to access information from anywhere within a wirelessly covered area using portable wireless enabled devices. We have presented our selection choices in order to deal with the hardware specific problems and described our policy based approach for access control enforcement. For interoperability issues we have selected the HL7 standard for codification of HL7 messages; in order to ensure confidentiality of messages we have used combination of public and shared key encryption techniques. We have also tested the validity of our approach by performing simulations in which we have estimated the number of queries and the resource consumption measuring the ability of our platform to respond.

We intend to apply our scenario to a wider extent and to measure our platform's performance through extensive testing and recording of efficiency parameters.

References

- [1] *The Pamvotis WLAN Simulator*, Information available online at www.pamvotis.org.
- [2] The Transport Layer Security Protocol Version 1.1, IETF, RFC 4346.
- [3] Wireless Medicenter. <http://www.wirelessmedicenter.com/mc/glance.cfm>
- [4] D. Brazier, Alpha Bravo Charlie Ltd. The m-care project. <http://www.m-care.co.uk/tech.html>
- [5] A. Malatras, G. Pavlou, P. Belsis, S. Gritzalis, C. Skourlas, I. Chalaris, "Deploying Pervasive Secure Knowledge Management Infrastructures", *International Journal of Pervasive Computing and Communications*, Vol. 1, No. 4, pp. 265-276, 2005, Troubador Publishing
- [6] Gritzalis S., Belsis P., Katsikas S. "Interconnecting autonomous medical domains: a security perspective" *IEEE EMB Magazine*, vol. 26, number 5, pp. 23-28, Sept.-Oct 2007, IEEE press
- [7] Gatzoulis S., Iakovidis I., "Wearable and Portable e-health systems *IEEE EMB Magazine*, vol. 26, number 5, pp. 51-55 Sept.-Oct 2007, IEEE press
- [8] A. Choudhri, L. Kagal, A. Joshi, T. Finin, and Y. Yesha, PatientService: Electronic Patient Record Redaction and Delivery in Pervasive Environments, Fifth International Workshop on Enterprise Networking and Computing in Healthcare Industry (Healthcom 2003), Santa Monica, June 2003
- [9] <http://www.pervasivehealthcare.dk/projects/index.html>
- [10] <http://www.eecs.harvard.edu/~mdw/proj/codeblue/>

Watermarking Medical Images with Anonymous Patient Identification to Verify Authenticity

Gouenou COATRIEUX^a, Catherine QUANTIN^b, Julien MONTAGNER^a, Maniane FASSA^b, François-André ALLAERT^c, Christian ROUX^a

a. Inserm U650, LaTIM; GET ENST Bretagne, Dpt. ITI

b. Dpt. of Biostatistics & Medical Informatics, Inserm U866, CHU de Dijon

c. Dpt. of Epidemiology and Biostatistics, Mc Gill University, Montreal Canada

Abstract: When dealing with medical image management, there is a need to ensure information authenticity and dependability. Being able to verify the information belongs to the correct patient and is issued from the right source is a major concern. Verification can help to reduce the risk of errors when identifying documents in daily practice or when sending a patient's Electronic Health Record. At the same time, patient privacy issues may appear during the verification process when the verifier accesses patient data without appropriate authorization. In this paper we discuss the combination of watermarking with different identifiers ranging from DICOM standard UID to an Anonymous European Patient Identifier in order to improve medical image protection in terms of authenticity and maintainability.

Keywords: Watermarking, Medical Imaging, Unique Patient Identifier, Authenticity, maintainability.

Introduction

The evolution of medical information systems, supported by advances in information technology, enables information to be shared between distant health professionals and manipulated and managed more easily. However, at the same time, more attention should be paid to information protection. For example, though the control of access to information has become tighter, when access is given, it is still difficult to guarantee that the information concerning a particular patient remains gathered in one place.

When dealing with information protection, we must distinguish between security and dependability. Security can be defined in terms of confidentiality, availability, integrity and authenticity [1]. In this paper, our interest concerns first of all authenticity; that is, providing proof that the information belongs to the correct patient and is issued from the right source. As we will discuss later, authenticity requires the creation of a code to identify one document uniquely and to establish a link between one document and one patient. This authentication code needs to accompany the

information it is associated with. Dependability mostly concerns the computing system. It can be described as a composite of availability (of a service), reliability (continuity of service) and maintainability (ability to undergo modifications and repairs). In this paper, we will discuss how maintainability with regard to medical images can be achieved using authenticity mechanisms.

Recently, watermarking has been proposed for medical information protection. Even though most of the work on watermarking has concerned medical images in order to verify image integrity or improve confidentiality [2], watermarking also provides a new way to share data. Basically, watermarking is defined as the invisible embedding or insertion of a message in a host document, an image, for example. In that way, watermarking is similar to steganography which means hidden (“stegano”) writing (“graphy”). However, contrary to steganography, with watermarking the dissimulated message is related to the host document and the presence of the message in the host is known to the users. As we will show later in this paper, watermarking makes it possible to introduce new security and management layers much closer to the host data: at the signal level. To our knowledge, very few studies have been devoted to authenticity control of medical images. For this purpose, we discuss in this paper ways to combine watermarking with different authenticity codes ranging from the UID of DICOM [3] to the European Identification Number introduced by Quantin *et al.* in [4].

1. Watermarking in healthcare

In this section we recall the relevance of watermarking as a complementary security mechanism for medical data within medical information systems.

1.1. Embedding data in host image data

A general chain of watermarking is depicted in figure 1. At the embedding stage, the message (stego-message) is inserted by modifying the host document in an “imperceptible” way. Such a host can be a signal, an image, a video, a text as a data base. “Imperceptible” means that the watermarked document can be used instead of the original document without interference.

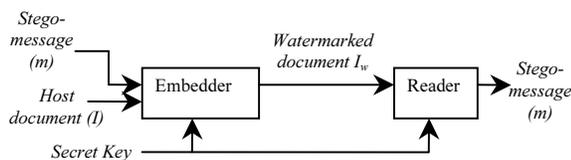


Figure 1. A general chain of watermarking.

Applied to an image; embedding consists in slightly modifying its pixel gray level values to insert the message. Two approaches are usually distinguished. A first set of methods, spread-spectrum-like methods, embeds one bit of the stego-message by adding a random sequence to some pixels of the image. This random sequence of gray values, which can be called a pattern, is derived from the secret watermarking key. For example, the key can be the seed of a pseudo random sequence generator. As a consequence, each bit of the message corresponds to the modification of several pixels in the image. The insertion can be viewed as the addition of a signal, a watermark w , to the image I . The watermark w corresponds to the set of random sequences, or patterns,

that encode the stego-message. At the reception of the image, extraction of each bit of message relies on the detection of each modulated random sequence in the watermarked image I_w through a correlation measurement. To improve imperceptibility of the watermark, “psychovisual masking” is used to determine maximum amplitude of the watermark that can be performed before this watermark becomes visible.

In the second type of method (such as Quantization Index Modulation [5]), an element from a dictionary is substituted for the original information. For example, one can substitute the least significant bit of the image with those of the stego-message. In this case, the dictionary creates the correspondence between the bits of the message and the parity of the gray levels (ex.: $255 \rightarrow 1$, $254 \rightarrow 0$, $253 \rightarrow 1$, ...). One simply has to read the image and interpret the observed gray values using the dictionary to decode the message. For more details about watermarking, the reader may refer to [5].

1.2. Applications in healthcare

1.2.1. Methods for watermarking medical images

Because modifying gray levels of a medical image may interfere with its interpretation and consequently with the diagnosis, specific methods have been proposed. These methods are based on the same principles as the methods previously described. One class of methods is based on reversible or lossless watermarking schemes. Once the embedded message has been interpreted, the watermark can be completely removed from the image, thus enabling the original image to be retrieved. Figure 2(a-c) gives an illustration of such a method [6]. In another method, an unimportant area of the image, the black background, for example, is watermarked. Such an approach leaves the information of interest for the diagnosis intact.

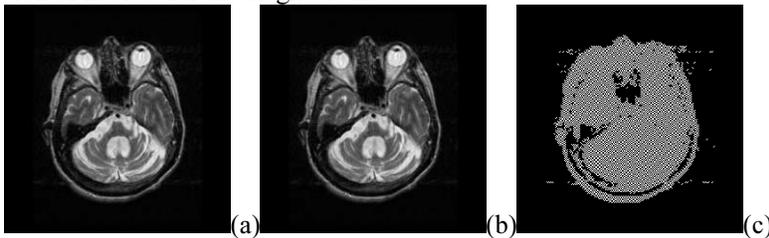


Figure 2. Illustration of the reversible watermarking method [6] used in a Magnetic Resonance Image of the Head (256x256 pixels, encoded on 12 bits), (a) Original image, (b) Reversible watermarked image (c) Signal of difference, it is the watermark w whose amplitude equals ± 1 or 0.

1.2.2. Interest in medical imaging

In medical imaging, it has been shown that watermarking can improve data protection and content enrichment. The insertion of meta-data facilitates data management [7][8]. For example, the medical knowledge illustrated by the content of one image can be summarized in a “knowledge digest” and shared with the image attached to its pixel values [8]. For medical image protection most applications have been devoted to integrity control and confidentiality. Regarding confidentiality, it is often considered that embedding makes it more difficult for unauthorized persons to gain access to the information [9]. In fact, it is more difficult to gain access to the embedded message than to an ancillary message. Integrity control can be achieved in different manners [2]. One simple way is to embed a digital signature of the image in the image itself. The

verification process will extract the embedded signature and compare it to the recomputed one. Any differences between the two signatures will state loss of image integrity. To our knowledge, few applications cover the issue of authenticity. However, watermarking could be of great help not only to verify authenticity but also to improve the maintainability of medical data (see section 3.1).

2. Watermarking for authenticity and maintainability of medical images

2.1. Authenticity and maintainability through watermarking

As defined, authenticity is based on proof that the information belongs to one patient and has been issued from the right source. This proof corresponds to an authenticity code (AC) associated with one image. The authenticity of an image can be verified in different ways with watermarking. If the AC of one image is known a priori (for example stored in the header of the image file) one approach is to verify the AC validity. To achieve this, one can embed in the image the sequence of bits which corresponds to its AC. As illustrated in figure 3a, the verification process will extract the AC and compare it with the one contained in the header. One constraint to be considered here is that in order to exploit the watermarking method it must be possible to insert the binary representation of the AC in the image. An alternative to this scheme is illustrated in figure 3b where the AC is used to generate the watermark signal (a random sequence, see section 2.1) which is then added to the image. In this case, authenticity verification relies on the detection of the watermark in the watermarked image. If the AC is not known a priori, this alternative requires testing all possible AC values and retaining the AC which provides the greatest correlation measurement, which is much more complex than with the first approach.

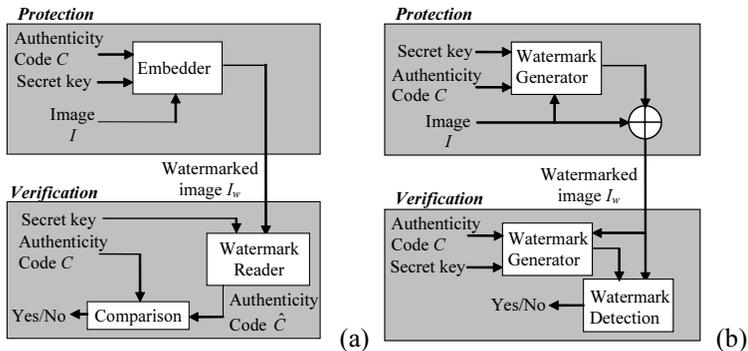


Figure 3. Verifying the authenticity of an image (a) the binary representation of the AC is embedded in the image (b) a watermark derived from the AC is added to the image.

As it becomes possible to retrieve the AC from the signal itself, watermarking can also be used to repair the link between one image, its origin and the patient it belongs to. This situation may occur after a change of the image file format, for example, when the original file header information is lost or altered. Hence, the embedded message simply has to be extracted from the image in order to recover the image's AC. In this way, it can help ensure system maintainability. However, the method should guarantee that the embedded AC can be retrieved exactly, which requires a watermark that is sufficiently robust to resist possible image alterations. The question of confidentiality

arises when verifying that the information belongs to the correct patient. How can we ensure privacy? This question has to be considered when structuring the AC.

2.2. Codes for authenticity

For images, an authenticity code is the combination of image and patient identifiers.

2.2.1. DICOM Image identifiers codes

DICOM (Digital Imaging and Communications in Medicine) is the standard of reference for medical images. This standard is developed by the American College of Radiology and the National Electrical Manufacturers Association, in liaison with other standardization organizations such as the CEN TC251 in Europe and the ANSI in the USA. DICOM technically guarantees data confidentiality, authentication of data origin, data integrity and digital signature through its international standards [3]. DICOM makes use of Unique Identifiers (UIDs) to uniquely identify DICOM objects such as images. These UIDs are based on the OSI object Identification as defined by the ISO 8824 standard. One UID, which can be defined privately, is constituted of two parts: <org.root>.<suffix>, each composed of a number of numeric components. The prefix <org.root> identifies an organization and is issued by a registration authority (ex. USA ANSI). The <suffix> is defined by the organization itself which has to guarantee uniqueness of the <suffix>. Consequently, and as stated by the standard, a UID only ensures uniqueness of one DICOM object. It cannot be parsed as it does not contain any semantics. As it is, such a UID is useful in the unique identification of an image. However, the issuing organization must not include information about the patient in the suffix as this data could jeopardize privacy once the suffix structure is known.

2.2.2. Patient identifiers (Id)

Several methods for patient identification have been developed. For example, DICOM proposes a method which seems complete and exhaustive, reliable and accurate. This method is based on a "Patient Module" that contains patient-related data such as: Patient's Name, Birth and Gender, Mother's Birth Name, Country and Region of Residence, Ethnic Group, Patient's Religious Preference and so on. Concerns related to DICOM data confidentiality with regard to the sensitive nature of the data can be raised. For example, the French authority for personal data protection has forbidden the communication of information such as Patient's Ethnic Group and Patient's Religion. A solution would be to render anonymous such patient information. However, there is no guarantee that this solution would be authorized by French authorities. To overcome this issue, and given that today there is no international harmonisation context for the patient identification [10], our view is to propose another patient identification method. In Europe, patient identification methods vary from country to country. Most of the North European Countries (Finland, Denmark, Luxembourg,...) use the national identity number for health purposes. In some countries, a specific national patient identification number is used, like in the United Kingdom, or planned like in the Netherlands and Ireland. In southern European countries, patient identification is based on regional specific patient identifiers. France and Belgium are developing a project related to specific healthcare national patient identifiers rendering anonymous the social security number. To guarantee interoperability of these different patient Id, we suggest keeping the national health numbers and combining them with an anonymized "pivot" Id, such as a family-based identifier referring to family medical records [4]. A

pivot Id ensures the link to the identifiers of other countries. Identifier calculation makes use of cryptographic hash function to ensure anonymity applied to a family-based identifier which is composed of nine key variables (last name, first name and date of birth of the patient, the patient's mother and the patient's father). The reader may refer to [10] for more details. This system has been validated by the French authority for personal data protection and patented (see international patent n° 11/683,003). The efficiency and accuracy of the method we propose for authenticity verification of images relies on: the incorporation of this anonymous patient identifier with the DICOM UID into an Authentication Code and, integration of the paired up identifiers into the image with watermarking. Patient information confidentiality will be ensured because the identifier is truly anonymous. This new method may allow medical image managers to gather the data of the same patient everywhere, anytime without knowing the true identity of that patient. With regard to management, utilization and secure access, it is very difficult to gather scattered data of the same patient. Our method thus provides a solution to these issues, while ensuring privacy.

3. Conclusion

Access to or sharing of an isolated medical document requires that the document can be identified. Watermarking can be used to provide proof of the authenticity of medical images, that is to say that the medical information belongs to one patient and has been issued from the right source. The possibility of inserting a watermark in a document to identify the patient, without interfering with the document's usefulness, will be a real step forward if the paradox of cryptic patient identification can be solved (anonymous for un-authorized users and accessible and available for those who are authorized). The quality of authentication also depends on the codes used. The DICOM proposal appears to be one of the best methods for the identification of the image and its source, but the user has access to the identity of the patient. The sensitive nature of some patient information obliges us to develop alternative methods. Our proposal combines an anonymized pivot number identifier with national patient identifiers so as to guarantee privacy and interoperability. This method may also provide a solution to the problem of the identification of lost medical documents.

References

- [1] G. Coatrieux, H. Maître, B. Sankur, Y. Rolland, R. Collorec, Relevance of watermarking in medical imaging, ITAB00, Arlington, USA, Nov. 2000.
- [2] G. Coatrieux, L. Lecornu, B. Sankur, C. Roux, A review of Image watermarking applications in healthcare, EMBC06, New York, USA, Sept. 2006.
- [3] Parts of the DICOM standard available at: <http://medical.nema.org/>
- [4] C. Quantin, F.A. Allaert, B. Gouyon, O. Cohen, Proposal for the creation of a European healthcare identifier, MIE05, Geneva, Switzerland; Stud. Health Technol. Inform., 116:949–954, Aug. 2005.
- [5] P. Moulin, R. Koetter, Data-hiding codes, Proc. of the IEEE, 93:2083–2126, 2005.
- [6] G. Coatrieux, M. Lamard, W. Daccache, J. Puentes, C. Roux, A low distortion and reversible watermark: application to angiographic images of the retina, EMBC05, Shanghai, China, Sept. 2005.
- [7] S. Cheng, Q. Wu, K.R. Castleman, Non-ubiquitous digital watermarking for record indexing and integrity protection of medical images, ICIP05, Genoa, Italy, vol. 2, Sept. 2005.
- [8] C. Le Guillou, G. Coatrieux, J.-M. Cauvin, L. Lecornu, C. Roux, Enhancing shared medical image functionalities with image knowledge digest and watermarking, ITAB06, Ioannina, Greece, Oct. 2006.
- [9] R. Acharya, U.C. Niranjan, S.S. Iyengar, N. Kannathal, L.C. Min, Simultaneous storage of patient information with medical images in the frequency domain, Comput. Meth. Prog. Bio. 76:13–19, 2004.
- [10] C. Quantin, O. Cohen, B. Riandey, F.-A. Allaert, Unique patient concept: a key choice for European epidemiology, Int. J. Med. Inform., 76:419–426, 2007.

Empowerment of Health Professionals: How High Level Security Education Can Raise Awareness and Confidence

Matthias HERBST ^{a,1}, Christoph BUSCH ^b, Peter PHAROW ^c and Bernd BLOBEL ^c

^a *Dermatologist, QSD, Darmstadt, Germany*

^b *Fraunhofer IGD, Darmstadt, Germany*

^c *eHealth Competence Center, Regensburg University Hospital, Germany*

Abstract. Setting up networks among physicians and other health professionals in virtually any medical discipline is an important part of establishing eHealth worldwide. Medical research strategies nowadays advance diagnostic and therapeutic knowledge and guidelines allowing patients to benefits. Patient data and samples are among the most sensitive information and must carefully be protected according to rules of ethics and professional discretion as well as national and international privacy legislation. A lot has been said about “patient involvement, patient empowerment”. What about health professionals? How can they be involved and empowered to address the paradigm shift towards a personalized health service provision? Information and communication technology (ICT), medical devices, and software applications are not among the topics health professionals typically deal with while being theoretically and practically trained to diagnose diseases and treat patients. An ICT-based training and information provision is required to update the ICT skills of health professionals. The German CAST association provides such an information platform where health professionals attend applied computer security education events. This article aims at describing how ICT and security education is provided to health professionals, and how these training courses are designed, structured, performed, and assessed.

Keywords: ICT, security, awareness, education, training, health professionals

Introduction

Information and communication technology (ICT) is the backbone of companies, research institutions, and the public sector structures. ICT infrastructure is an essential part of the organizational infrastructure as such. More and more, ICT gets important for small business and privateers regardless whether they typically search for reliable information on the Internet, write or receive emails, or use online shops. Any violence of, or attack to, such an ICT may cause unauthorized access to, and modification of, sensitive data, loss of data, loss of availability of services, and loss of respective perform-

¹ Corresponding Author: Dr. Matthias Herbst, Dermatologist, South Hessian Dermatological Association for Quality Management (QSD), Merckhaus, Rheinstrasse 7-9, 64283 Darmstadt, Germany. Email: matthiasherbst@web.de

ance. Such threats must be analyzed, risks must be validated, countermeasures must be implemented in order to minimize the probability that such situations occur. It's not only a financial loss; it's often a loss of reliability, trustworthiness, and creditability.

ICT security is thus of high and even growing importance in virtually any society sector or business domain. But security itself can only be guaranteed in case the responsible parties (user, administrator, chief information officer, IT security officer, data protection and privacy ombudsperson, etc.) are aware of threats and risks, and are keen on identifying and applying countermeasures. This requires awareness, security education, and permanent security training. A slogan that really gets it to the point is "Security is not a product - it's a process" [1]. It means one can't buy the perfect security solution just from the stock. Security is a permanent, dynamic process rather than being a static product. Security policies and procedures need to dynamically respond to immediate changes in security infrastructures as such (e.g. new medical devices, applications, updates, upgrades, patches, and migrations) and a related different threat and risk potential caused by the modified ICT environment situation. Advanced education and training is a must to keep even security-aware personnel permanently updated [2].

1. General ICT Security Experience and Expertise Requirements

Healthcare and welfare around the globe in developed countries, developing countries, and countries in transition is turning towards the paradigm of shared care providing an integrated care approach. The underlying paradigm change is bound to an extended communication and cooperation network between different specialized healthcare providers involved in patient's care. Challenges of increasing quality and related efficiency and efficacy requirements in the respective domains need to be met not only by local and regional providers but increasingly by existing and emerging national and even European networks of healthcare establishments and health professionals.

Advanced health networks including those bridging between persons have to be established and managed from different viewpoints. Legislation and administration needs to be addressed. Another important aspect deals with the domain of technologies with especially the industry and the responsible standards developing organizations (SDO) being in the driver's seat. Last but not least, the societies shall be addressed by the European Commission, the Member States, and the different stakeholder communities. Special attention needs to be paid to ethics, history, and culture. So networking among stakeholders within and between domains is of growing importance for establishing a stable collaboration and cooperation across national and domain boundaries.

Security requirements in the medical area are technically not that different from other application domains. Apart from a very demanding and dynamic privilege management and access control policy, medical solutions (hospital information systems, public health information systems, radiology information systems, laboratory systems, GP office software, and many other applications) base their security solutions on proper authentication (identification and verification), identity management, confidentiality, integrity, authenticity, availability, and accountability [3], [4].

2. ICT Security Experience and Expertise among Medical Professionals

ICT knowledge and awareness for ICT usage needs to be ranked high in medical professionals' daily routine. First approaches of a permanent working group called "Quality and Security in the dermatological office" in Germany to improve ICT security in practice were made back in 1993. A list of possible pitfalls was circulated among medical professionals immediately after an accidental distribution of confidential patient data by an ICT service engineer. The following investigation of the case clearly showed that no written service and confidential policy agreements were common in practice up to this time. As a first step, the working group established a mailbox system to distribute important findings regionally to dermatologists. In 1996, this service was integrated into a First Class Server based solution with high level security. This did lead to an intensive discussion on quality and security objectives in medical professionals' work. In 1998, after having started an intensive cooperation with the German Fraunhofer Institute for Computer Graphics (IGD) and its security experts, the gained experience was integrated as valuable block of knowledge into the in the meantime established Competence Center for Applied Security Technology (CAST) [5].

2.1. Networking and Security Expertise

As of today, most physicians are not really familiar with modern and advanced security technology. The physician's Internet Service Provider (ISP) is considered responsible for all related security measure. Nevertheless there exists almost no educational proof and no quality test (service certification, quality labeling) to assure that a certain level of experience is guaranteed to be achieved by that service provider. The decentralized location of medical treatment makes it difficult to provide a defined quality region-wide. The forthcoming implementation of Electronic Health Record (EHR) systems and the frequent exchange of sensitive medical data via the open and policy-free Internet make strong and high-level quality and security control measures necessary. Networks of medical professionals start making use of advanced ICT solutions [6].

Telemedicine is more and more being accepted by the medical community worldwide. The secure and reliable exchange of data and images of a patient whose life depends on the correct diagnosis at the right place right in time makes the strongest available security efforts inevitable. The screening of moles in high risk patients, e.g., is a live-long process which makes consequent and secure data storage and exchange absolutely necessary. As a well-known procedure now being applied in medicine as well - the central storage of important data in high security server plants- has to seriously be discussed with all its legal, ethical, and technical advantages and drawbacks. The reliable and liable exchange of administrative and medical (clinical) patient data becomes indeed effective and more reliable by using the much faster DSL technology. Disaster scenarios and emergency concepts are well known in military and civil services all over the world, and have increasingly to be applied in the field of medicine.

Assessment studies and reports on domain-independent security training for domain specialists (like medical professionals) have clearly shown the need for specific education and training concepts comprehensively addressing the domain needs [7], [8].

2.2. ICT Expertise

Permanent training and further education is essential for medical professionals to be aware of both legal requirements and technical solutions for advanced ICT security. To address these aspects, the current eLearning initiative in medicine in general and especially in dermatology was launched about five years ago. Meanwhile, the first course of tele-dermatology concerning the diagnosis of early stages of malignant melanoma is under way. In cooperation with University of Brisbane, Australia [7] and University of Graz, Austria [8], dermatologists from different countries world-wide train the exact and early cancer detection in a course of high level web based education. The Armed Forces Institute of Pathology, USA, uses the Internet for exchanging medical expertise with, and between doctors from all over the world for years [9]. Their seminars are highlights in international pathology diagnosis and learning. International networks and discussion groups are supported by services such as SKYPE [10]. Nevertheless, the contact between patients and their doctors is ameliorated, e.g., by using a calendar of symptoms, by exchanging images and digital photos demonstrating the healing process after surgery, or simply by the coordination of visits in the physician's office. All these aspects need to be addressed while establishing advanced ICT in the medical domain.

Security is one of the basic principles of modern ICT in medicine. It is a part of the obligatory quality management bases for medical institutions. Therefore all parts of such initiatives should undergo a deep and intensive testing procedure before being rolled out nation-wide. Experience and pitfalls have to seriously be evaluated. The influence of ICT industries and their need to foster effective processes supporting economical interests must not lead to a speedy implementation of unreliable or even insecure technology. Especially an independent forum like CAST is able to independently evaluate best practice procedures for ICT security in medical practices and hospitals to avoid health system collapses by hazardously applied techniques and technologies.

3. The CAST Concept

The Darmstadt-based Competence Center for Applied Security Technology (CAST) was established in 1999 [5]. It aims at bringing security expertise to user communities. One way doing so is a series of workshops. These workshops address a specific security area or topic. Experts introduce recently researched security developments and present their chances and challenges. Intensive discussions with the audience validate requirements and conclusions. In collaboration with the Darmstadt Center for ICT Security (DZI) [11] at the Darmstadt Technical University (TUD) [12], CAST is able to offer a complementary a high-class scientific education and training stream. Within four semesters, ICT employees are entitled to undergo an advanced security course finally awarded with an academic certificate on "ICT Security" after having successfully completed a certain number of academic modules and workshops. CAST and DZI offer this certificate in a joint effort including regular courses at TUD. Specifically designed and developed materials even increase the knowledge on ICT security as course participants are entitled to specific additional training and practice courses. CAST workshops typically provide up-to-date information on hot ICT security topics.

The CAST workshops can meanwhile be considered a backbone of German ICT security education. Workshops typically base on knowledge provided in tutorials, and offer state-of-the art experts knowledge in addressed field of interest. In terms of a

Public Key Infrastructure (PKI), e.g., the tutorial offers basic information about public key (asymmetric) encryption, electronic (digital) signatures, policies, establishment of Trusted Third Party (TTP) services, and advantages / drawbacks of existing and emerging PKI solutions. The workshop itself highlights current technical developments and software / hardware solutions in this domain. Tutorials and related workshops shall be seen as a package thus providing real up-to-date information on the addressed topic.

The list of CAST workshop topics is quite comprehensive and covers application areas ranging from cryptography, biometrics, protocols, privacy, and data protection to security engineering, network security, public key infrastructures, and mobile security. Workshops on security levels and related evaluation as well as on identifying specific security threats complete the spectrum. Complementary workshops address specific business domains like, e.g., security management, electronic payment, identity management, electronic procurement, email, cyber crime, and -last but not least- the medical domain with a yearly MED-CAST workshop. Participants are, e.g., ICT security experts, ICT administrators, ICT managers and decision makers, politicians, end users, SMEs, medical professionals, health professionals, and data protection ombudspersons.

4. The MED-CAST Concept

Medical and health professionals do not spend much time in identifying security threats and risks. Thus, one of the CAST workshops aims at inviting medical professionals and ICT specialists involved in healthcare application development for about 6 years now [5]. MED-CAST addresses different stakeholder groups like people working in healthcare, eHealth experts, ICT employees in healthcare, developers and suppliers of medical devices, security experts, medical and technical safety experts, and many other groups. The number of domain participants is promising (2002: 49, 2003: 61, 2004: 55, 2005: 52, 2006: 32 attendees). The workshop itself consists of 8 to 10 presentations with discussions addressing different aspects of the topic, followed by a panel discussion with an active audience interaction. The feedback is very positive in general.

CAST performs for all workshops an extensive assessment to address awareness and acceptance as well as satisfaction of the attendees and ideas for future events. MED-CAST ranges among the best-assessed workshops. Presentations are evaluated regarding their scientific / educational values and their clarity in presentation. Despite of their basically limited ICT security knowledge, they rated most of the presenters quite high and appreciated workshops as such (for rating details see Figures 1 and 2).

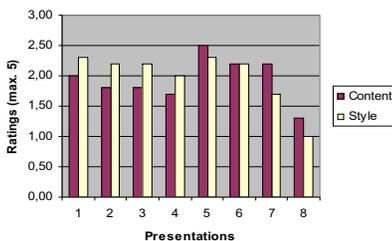


Figure 1. MED-CAST 2004 Evaluation

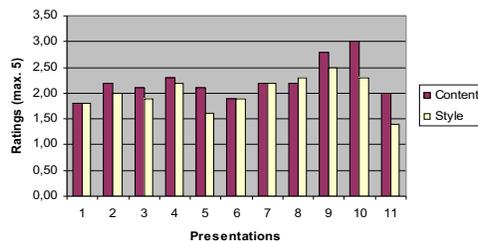


Figure 2. MED-CAST 2006 Evaluation

MED-CAST started back in 2002 with an introductory workshop on ICT security aspects in healthcare and welfare. The 2003 workshop addressed infrastructure security

for the medical domain whereas the events in 2004 and 2005 highlighted the importance of secure solutions for applying health cards and health records in Germany and beyond. The 2006 workshop tackled challenges and solutions for secure medical multimedia data storage and exchange including images, audio and video streams. In 2007, the organizers prepared, for the first time at least partly in English language, an event on national security infrastructure experience in Germany and its neighbor countries. The 2008 topic (on April 17th) is related to ICT security in medical imaging.

Over the year educational measures and training courses have increasingly been considered important for medical professionals with speakers from different domains and viewpoints highlighting various exciting new topics of applied security technology in the medical domain. Several field studies concerning the secure transport of visual information were performed meanwhile in cooperation with CAST and IGD.

5. Conclusion

Medical professionals typically have many different tasks others that dealing with ICT security. It's the treatment of patients that count. So in terms of advanced ICT solutions, they need service providers. But in order to validate and evaluate the quality of the services offered, even medical and health professionals need to have a basic understanding of security and safety in ICT. Threats and risks have at least to be known to them. Awareness of security and safety threats as well as confidence in applied security solutions leads to a higher acceptance level for advanced and secure communication and technical collaboration. The MED-CAST workshops held once a year in Darmstadt have been accepted by the expert community as a forum where speakers from different points of view highlight various exciting new themes and topics of applied security technology. Lots of users and scientists have visited the annual symposia meanwhile. MED-CAST is well established in the domain of health-related advanced applied security knowledge provision. The authors will therefore keep the concept of MED-CAST workshop series running. The 2008 topic is related to ICT security in medical imaging.

Acknowledgement

The authors are in debt to Fraunhofer IGD, CAST as well as their own organizations and partnering experts and associations for their continuous support and cooperation.

References

- [1] Bruce Schneier, "Security Is Not a Product; It's a Process". Online: <http://www.schneier.com/cryptogram-9912.html> (last accessed February 25th, 2008)
- [2] H. Baier, J. Buchmann, C. Busch: "Aus- und Weiterbildung in IT-Sicherheit" in Tagungsband 8. Deutscher IT-Sicherheitskongress des BSI 2003, SecuMedia Verlag, pages 179-190, 2003.
- [3] Matt Bishop, Computer Security – Art and Science, Addison-Wesley, 2003.
- [4] Pharow P, Blobel B: Von eHealth zu mHealth – die Sicherheitsanforderungen in der mobilen Welt. Proceedings of GMDS 2007. <http://www.med-ges-2007.de/> (in German).
- [5] CAST Online: <http://www.cast-forum.de> (last accessed February 25th, 2008).
- [6] Kompetenznetz Angeborene Herzfehler. Online: <http://www.kompetenznetz-ahf.de/> (last accessed February 25th, 2008).
- [7] Peter Hoyer, University Brisbane, AU, status of 2007
- [8] Helmut Kerl, Medizinische Universität Graz, status of 2007
- [9] The Armed Forces Institute of Pathology, status of 2007
- [10] SKYPE Online: www.skype.com (last accessed February 25th, 2008)
- [11] Center for ICT Security Online: <http://www.dzi.tu-darmstadt.de> (last accessed February 25th, 2008)
- [12] TU Darmstadt, Online: <http://www.tu-darmstadt.de/weiterbildung/> (last accessed February 25th, 2008)

Context-Aware Access Control for Pervasive Access to Process-Based Healthcare Systems

Vassiliki KOUFI and George VASSILACOPOULOS

Department of Digital Systems, University of Piraeus, Piraeus 185 34, Greece

Abstract. Healthcare is an increasingly collaborative enterprise involving a broad range of healthcare services provided by many individuals and organizations. Grid technology has been widely recognized as a means for integrating disparate computing resources in the healthcare field. Moreover, Grid portal applications can be developed on a wireless and mobile infrastructure to execute healthcare processes which, in turn, can provide remote access to Grid database services. Such an environment provides ubiquitous and pervasive access to integrated healthcare services at the point of care, thus improving healthcare quality. In such environments, the ability to provide an effective access control mechanism that meets the requirement of the least privilege principle is essential. Adherence to the least privilege principle requires continuous adjustments of user permissions in order to adapt to the current situation. This paper presents a context-aware access control mechanism for HDGPortal, a Grid portal application which provides access to workflow-based healthcare processes using wireless Personal Digital Assistants. The proposed mechanism builds upon and enhances security mechanisms provided by the Grid Security Infrastructure. It provides tight, just-in-time permissions so that authorized users get access to specific objects according to the current context. These permissions are subject to continuous adjustments triggered by the changing context. Thus, the risk of compromising information integrity during task executions is reduced.

Keywords: Grid portal application; process; security; role-based access control (RBAC); managing change.

Introduction

Healthcare delivery is a highly complex process involving many individuals and organizations [1]. Providing readily access to integrated healthcare information at the point of care remotely requires a system architecture that enables collaboration and coordination among healthcare services and facilitates the mobility of healthcare professionals. To this end, a prototype Grid portal application, namely HDGPortal, has been developed that provides pervasive access to process-based healthcare systems [1]. This paper focuses on the security mechanism incorporated into the HDGPortal.

HDGPortal is a portal application that is used to run healthcare processes implemented in the Business Process Execution Language (BPEL). These processes invoke Grid database services, while on execution, in order to provide integrated access to healthcare information scattered on a Grid computing infrastructure. As tight matching of permissions to actual usage and need is essential in healthcare applications, one important security requirement in HDGPortal is adherence to the *least privilege principle*. Its enforcement requires continuous adjustments of user permissions to ensure that users assume the minimum sets of permissions required for the execution of each task of a healthcare process selected. To this end, changes in contextual information should be sensed during task executions and relevant actions should be fired in response to these changes.

In the last few years, there has been a trend towards using BPEL for Grid service composition [2]. Security aspects, such as authentication and access control, are not standardized through BPEL, but are left to the implementation of BPEL compliant process engine [3,4,5]. In turn, grid middleware, namely Open Grid Services Architecture - Data Access and Integration (OGSA-DAI) [6], that facilitates data federation and distributed query processing through the use of grid database services provides relatively simple and static mechanisms regarding authorization and access control [7]. Several studies have been conducted regarding both the enforcement of access control in BPEL [3,4,5,8] and the enhancement of access control mechanisms used by Grid middleware services [7,9]. Most of these studies argue that both BPEL and Grid middleware services can benefit from incorporating properly enhanced role-based access control (RBAC) mechanisms [15]. However, they are not concerned with access control in the context of a pervasive process-based healthcare system build on a Grid infrastructure.

This paper is mainly concerned with situations, often occurred in healthcare delivery, where the information requested by an authorized user needs to be available when and where needed. Thus, a context-aware access control mechanism is proposed that incorporates the advantages of broad, role-based permission assignment and administration across object types, as in RBAC, and yet provides the flexibility for adjusting role permissions on individual objects during a BPEL process enactment according to the current context. During the execution of a workflow instance, changes in context information are sensed to adapt user permissions to the minimum required for completing a job. Relevant access control policies are enforced at both the BPEL task level and the Grid database service level.

1. Motivating Scenario

To illustrate the main principles of the security architecture incorporated into the HDGPortal, a sample integration project is described which is concerned with the automation of a cross-organizational healthcare process spanning a health district. Typically, a health district consists of one district general hospital (DGH) and a number of peripheral hospitals and health centers. As patient referrals are usually made among various healthcare providers within a district (e.g. for hospitalization, for outpatient consultation or for performing specialized medical procedures), there is a need to ensure authorized access to the tasks comprising the relevant healthcare

processes and, then, to patient information required through the execution of these tasks.

Suppose that a healthcare process is concerned with a radiological request issued by physicians on a ward round, for their patients. The radiological department

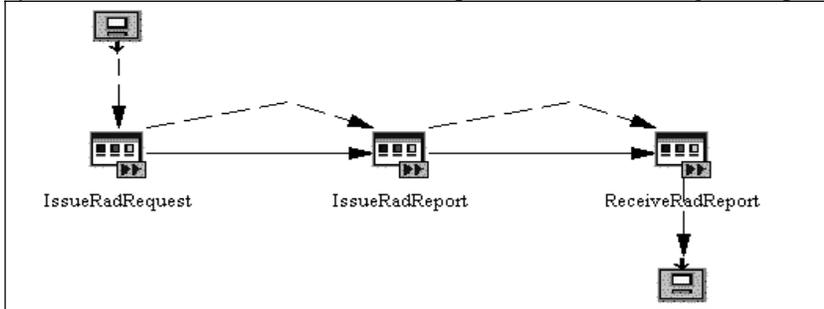


Figure 1. Radiological request process model using IBM WebSphere Workflow

receives each request, schedules the radiological procedures requested and sends a message to the requesting physician notifying him/her on the date and time scheduled for its performance. On performing the radiological procedure requested, the radiologist accesses the relevant part of the patient record and issues a radiological report, incorporating both the radiological images and the associated assessment, which is send to the requesting physician.

Figure 1 shows a high-level view of the healthcare process using IBM Websphere Workflow build-time tool [10]. In this business process the hospital organizational units involved are the clinical department and the radiology department of a hospital and the roles participating in the healthcare process are physician and radiologist.

From an authorization perspective, the healthcare process of Figure 1 surfaces several requirements with regard to task execution and Grid database services invocation. These requirements include the following:

- **Restricted task execution**

In certain circumstances the candidates for a task instance execution should be dynamically determined and be either a sub-group of the authorized users or only one, specific authorized user. For example, a request for performing a radiological procedure on a patient (e.g. CT or MRI), issued by a physician, should be routed only to the sub-group of on-duty radiologists who hold the relevant sub-specialty and the radiological report, issued by the radiologist, should be routed only to the requesting physician.

- **Restricted grid database service invocation**

Given that a role holder can execute a specific task, he/she should be allowed to exercise a dynamically determined set of permissions on certain data only which are accessible via the associated Grid database services. For example, during the execution of the “IssueRadRequest” task, the relevant Grid database services invoked should allow a physician to write and read patient record data and to issue (write, edit and send) radiological requests only for his/her patients while on duty and only within the hospital premises.

The above requirements suggest that certain permissions of the healthcare process participants depend on the process execution context. In particular, contextual information available at access time, such as proximity, location and time, can

influence the authorization decision that allows a user to perform a task and, given this permission, to invoke the associated Grid database services. This enables a more flexible and precise access control policy specification that satisfies the least privilege principle.

2. Access Control Architecture

Access control constitutes a major research issue in traditional and pervasive Grid environments [1]. Figure 2 shows a high-level view of the security architecture implemented in HDGPortal. This is described by a two-tier model consisted of a global access control service, residing on a server at the DGH site, and one local access control service, residing at each healthcare organization within the health district.

2.1. Access Control Mechanism

In the HDGPortal prototype, access control is provided at two levels: the BPEL task level and the Grid database service level [1]. Hence, a *middleware-based* access control mechanism has been developed which is employed to mediate between subjects (healthcare professionals) and objects (BPEL tasks and Grid database services) and to decide whether access of a given subject to given object should be permitted or denied by taking into account the current context. In particular, the Java Authentication and Authorization Service (JAAS) [11] was used for the development of an external to the BPEL engine access control service that regulates user access to tasks, and an external to OGSA-DAI, access control service that enhances its mechanism by adding context-awareness features.

The developed mechanism is *certificate-based* as it relies on Community Authorization Service (CAS) certificates issued to healthcare professionals by a CAS server residing at the DGH site. These certificates specify user-to-role assignments in the form of security assertions, expressed in Security Assertion Markup Language (SAML) [12,13]. CAS certificates accompany every request (either for task execution or Grid database service invocation) issued through HDGPortal. The roles used in the certificates are functional and, hence, they remain unchanged until the certificate expires as they are independent of the constraints held at the time of attempted access.

The mapping of the aforementioned roles to the relevant permissions is performed by means of access control policies expressed by using the RBAC profile of eXtensible Access Control Markup Language (XACML) [13]. These policies are specified at the site where the target object (task or Grid database service) resides (tasks are hosted on the BPEL engine at the DGH site and Grid database services are hosted on the web servers at the hospital sites) and assist in the derivation of the exact permissions a subject should acquire for performing a task. In particular, on issuing a request for a task execution, the roles contained in the CAS certificate accompanying the request are extracted and their relevant permissions regarding access to BPEL tasks are specified using a file storing the XACML policies [13]. This file resides on the same server of the DGH site with the BPEL engine. Then, during task execution, a request for invocation of the underlying Grid database services is issued which is accompanied by the same CAS certificate. The roles extracted from this certificate are

used in order to specify the relevant permissions regarding Grid database services using XACML policies stored in one file at each Grid node (i.e. healthcare

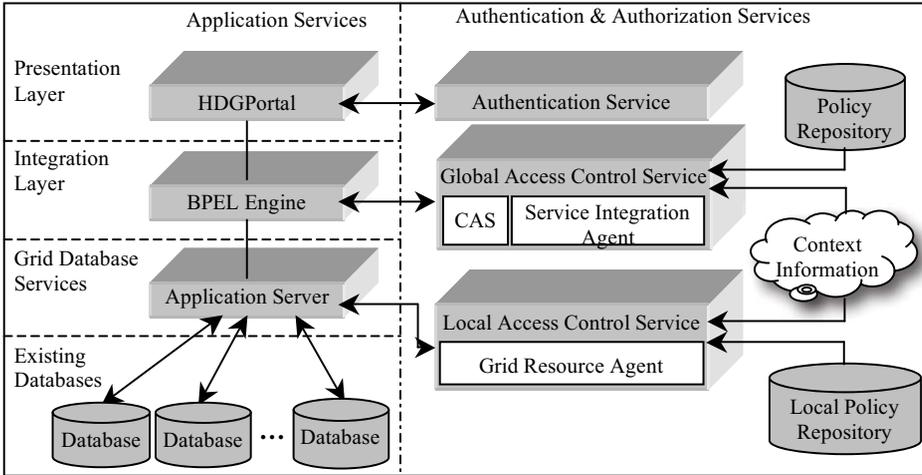


Figure 2. Security architecture in HDGPortal

organization) providing the portion of medical information requested. Permissions on both BPEL tasks and Grid database services are dynamically adapted by the constraints imposed by the current context.

2.2. Context Information Management

In the HDGPortal prototype, the contextual information is determined by a pre-defined set of attributes related to the user (e.g. user certificate, user/patient relationship), to the environment (e.g. client location and time of attempted access) and to the data resource provider, namely to the healthcare organization (e.g. local security policy). For example, the permissions of a physician using HDGPortal on his/her PDA, are adapted depending on his/her identity (included in the CAS certificate), location and time of access as well as the security policy of each healthcare organization where a portion of the requested information is stored.

Context information is collected by a Context Manager which has been implemented as a multi-agent system. Thus, the Context Manager consists of two kinds of agents, developed in JADE [14]:

- *Service Integration Agent*: it is hosted on a server at the DGH site and manages user permissions on BPEL tasks
- *Grid Resource Agent*: it is hosted on a server at the site of each healthcare organization participating in a healthcare process and manages user permissions on Grid database services.

Each agent uses middleware context collection services to monitor context and interacts with a state machine that maintains the permission subset of each role. The state machine consists of variables that encode state (permissions of each role), and events that transform its state. At the time of an attempted access (either to a task or to

a Grid database service), the relevant agent generates an event to trigger a transition of the state machine. Changes in user and environment context are sensed by both agents, whereas changes in resource context are sensed and dealt with by the Grid resource agent lying at each Grid node.

3. Concluding Remarks

Development of pervasive applications that provide readily access to integrated healthcare information at the point of care, introduces security risks especially with regard to authorization and access control. Hence, relevant mechanisms must be in place that can conveniently regulate user access to information while providing confidence that security policies are faithfully and consistently enforced within and across organizations residing in a health district. In particular, when adherence to the least privilege principle is considered a prominent feature of a system, the incorporated access control mechanism should provide tight, just-in-time permissions so that authorized users get access to specific objects subject to the current context. The access control mechanism presented in this paper meets the aforementioned requirements and is embedded into a Grid portal application, namely HDGPortal. In particular, the mechanism ensures authorized execution of BPEL tasks and invocation of relevant Grid database services in accordance with the current context. Thus, a tight matching of permissions to actual usage and need is ensured.

References

- [1] Koufi V, Vassilacopoulos G. HDGPortal: A Grid Portal Application for Pervasive Access to Process-Based Healthcare Systems, Proceedings of the 2nd International Conference in Pervasive Computing Technologies in Healthcare, 2008.
- [2] Emmerich W, Butchart B, Chen L, Wassermann B, Price S. Grid Service Orchestration Using the Business Process Execution Language (BPEL), *Journal of Grid Computing* (2006) 3: 283-304.
- [3] Mendling J, Strembeck M, Stermsek G, Neumann G. An Approach to Extract RBAC Models for BPEL4WS Processes, Proceedings of the 13th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, 2004.
- [4] Thomas J, Paci F, Bertino E, Eugster P. User Tasks and Access Control over Web Services, Proceedings of the 15th IEEE International Conference on Web Services, 2007.
- [5] Bertino E, Crampton J, Paci F. Access Control and Authorization Constraints for WS-BPEL, Proceedings of the IEEE International Conference on Web Services, 2006.
- [6] Open Grid Services Architecture - Data Access and Integration (OGSA-DAI), <http://www.ogsadai.org.uk/>.
- [7] Adamski M, Kulczewski M, Kurowski K, Nabrzyski J, Hume A. Security and Performance Enhancements to OGSA-DAI for Grid Data Virtualization, Concurrency and Computation.: Practice and Experience, 2007.
- [8] Dou W, Cheung SC, Chen G, Cai S. Certificate-Driven Grid Workflow Paradigm Based on Service Computing, *Lecture Notes in Computer Science* (2005) 3795: 155-160.
- [9] Power D, Slaymaker M, Politou E, Simpson A. A Secure Wrapper for OGSA-DAI, *Lecture Notes in Computer Science* (2005) 3470: 485-494.
- [10] IBM Corporation. IBM Websphere Workflow – Getting Started with Buildtime V. 3.6, 2005.
- [11] Java Authentication and Authorization Service, <http://java.sun.com/javase/6/docs/technotes/guides/security/jaas/JAASRefGuide.html>.
- [12] Pearlman L, Welch V, Foster I, Kesselman C, Tuecke S. A Community Authorization Service for Group Collaboration. Proceedings of the 3rd IEEE International Workshop on Policies for Distributed Systems and Networks, 2002.
- [13] OASIS Standards, <http://www.oasis-open.org/>
- [14] Java Agent Development Framework, <http://jade.tilab.com/>
- [15] National Institute of Standards and Technology (NIST) RBAC, <http://csrc.nist.gov/groups/SNS/rbac/>

Knowledge Management for the Protection of Information in Electronic Medical Records

Nathan LEA ^{a,1}, Stephen HAILES ^b, Tony AUSTIN ^a and Dipak KALRA ^a

^a *The Centre for Health Informatics and Multiprofessional Education, University College London, United Kingdom*

^b *Department of Computer Science, University College London*

Abstract: This paper describes foundational work investigating the protection requirements of sensitive medical information, which is being stored more routinely in repository systems for electronic medical records. These systems have increasingly powerful sharing capabilities at the point of clinical care, in medical research and for clinical and managerial audit. The potential for sharing raises concerns about the protection of individual patient privacy and challenges the duty of confidentiality by which medical practitioners are ethically and legally bound. By analysing the protection requirements and discussing the need to apply policy-based controls to discrete items of medical information in a record, this paper suggests that this is a problem for which existing privacy management solutions are not sufficient or appropriate to the protection requirements. It proposes that a knowledge management approach is required and it introduces a new framework based on the knowledge management techniques now being used to manage electronic medical record data. The background, existing work in this area, initial investigation methods, results to date and discussion are presented, and the paper is concluded with the authors' comments on the ramifications of the work.

Keywords: Security, legal issues, compliance, repository system for medical records, knowledge management, ethics, patient identification, distributed systems, sharing personal data, personal privacy, confidentiality

Introduction

Research into establishing shareable, clinically meaningful and accurate Electronic Medical Records (EMR) has been continuing for about fifteen years to improve the provision of information resources for effective medical care. This research has produced several examples of repository systems for medical records across the world [1] [2], which facilitate disease management, decision support and patient monitoring for millions of individuals. These repositories can also support medical research, where information about large numbers of patients is made available to help improve care provision. The sharing of larger quantities of data is made easier by the

¹ Corresponding Author: Nathan Lea, CHIME, University College London, 4th Floor, Holborn Union Building, Highgate Hill, London, N19 5LW, United Kingdom; E-mail: n.lea@chime.ucl.ac.uk

international standardisation and use of information models (from Health Level 7 [3], EN / ISO 13606 and *openEHR* [4], amongst others), and the use of Archetypes to “provide the common basis for ubiquitous presence of meaningful and computer-processable knowledge and information” [5].

There are legal and ethical responsibilities to protect this information: the rights of the subject of information are considered in law to be paramount, and this gives rise to constraints on how clinicians and researchers may behave with respect to the information so that no harm comes to individuals as a result of the increased accessibility. The constraints exist in the form of legislation and standards (summarised in Section 2), which inform institutional and research governance in the form of policies [6] [7] at the level of both research institutions and healthcare authorities: these policies may have different human interpretations and are often difficult to implement.

The security literature recommends the use of policy-based controls to manage the protection requirements (as per ISO 17799, discussed in section 2), but this requires interpretation in the same way as legislative controls. The use of roles-based access control (RBAC) is a keenly researched methodology for computable policy implementation, but the process of interpreting and operationalising each policy remains manual, loses much of the semantic sense when put into computable form (as Becker illustrates [8]), and relies on the use and configuration of generic software solutions. These are not holistic in scope, have insufficient structure, syntax and weak semantics, and are under-specified for individual classes of EMRs.

This paper proposes a knowledge management framework to meet the comprehensive needs of managing policy-based controls governing a repository system for medical records, and advocates that this framework uses similar techniques to those that manage EMR knowledge (i.e. – Archetypes). This approach is expected to complement existing solutions so that their manual configuration and risks of human fallibility in policy specification can be alleviated.

1. Methods Used for Initial Investigations

The methods for investigating this problem space have included literary reviews of: legislation (including the UK Data Protection Act 1998, Human Rights Act 1998, Freedom of Information Act 2000 and National Health Service Act 2006 (sections 251 and 252) as well as European Equivalents); informatics knowledge management techniques (including the standards work for information models, construction of EMR repositories [9] [10]); commentaries on the sharing of information and associated risks of centralising information [11]; and existing solutions (including the Ponder Policy Specification [12] and Cassandra [8] amongst others). Standards reviewed include Information Security Management ISO 17799; the draft ISO 27799 for healthcare information security management; and the draft ISO 22271 for constructing EMR repositories for research. The academic literature is being continuously reviewed for indications of issues that have arisen and lessons learned in the deployment of EMR repositories.

Formal and informal interviews have been conducted to support the literature reviews. Medical practitioners and researchers are particularly useful in helping to explore current working practice, levels of knowledge and awareness of information protection issues and difficulties raised by the legislative and governance requirements.

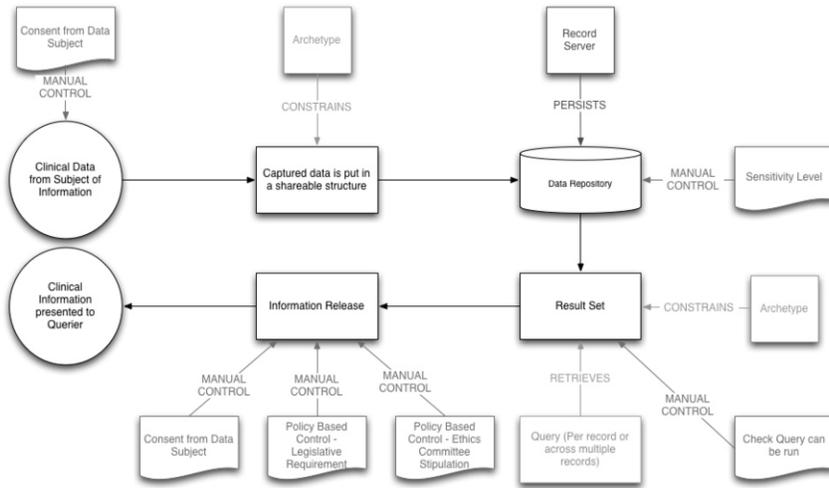


Figure 1. Information Flow and Management in EHR Supported Services

Wider discussions with legal experts, legislators and security experts have helped to refine the requirements of protection and policy setting.

There are often examples of the needs of the informatics community at more general conferences (particularly MedInfo 2007) and meetings that illustrate concepts for tooling and working practices. Further investigation of the protection requirements has been undertaken through the practical experience offered by the Clinical eScience Framework project [13]. The results of these reviews and discussions have been used to establish a set of preliminary requirements for the protection of sensitive data in EMR repositories. These requirements have been refined to model current working best practice in managing EMRs and proposed best practice.

2. Results

The reviews and discussions to date have yielded some preliminary conclusions about the process of protecting information stored in and shared by EMR repositories. Figure 1 illustrates current good practice in terms of managing the storage and retrieval of EMRs by Archetypes, assuming an EN 13606 or related scenario; the protection measures are included in the diagram and are labelled as ‘MANUAL CONTROL’.

The requirements for the protection measures show that there is a need for individual patient consent to use the information, but gaining such consent is not a scalable activity, especially in the case of medical research where thousands or millions of patients’ records can now be accessed. The de-identification of patient information, where identifying attributes such as names, dates of birth, hospital identity numbers and postcodes are removed, is considered by medical ethics practitioners and in law to be a reasonable means of assuring some level of confidentiality where consent cannot be (or is not) gained [14], but it is very hard to facilitate in practice [11]. Furthermore, complex patterns of roles and data-uses have started to emerge as the sharing of medical information in different contexts of use becomes easier.

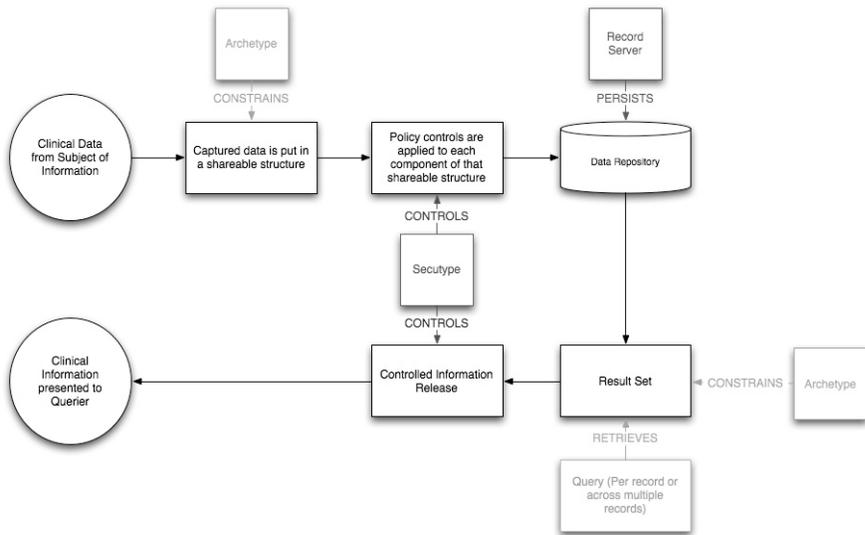


Figure 2. Application of knowledge management component for information security control - Secutypes

It is clear that security controls need to be applied to individual data items, and those controls need to be asserted based on details about who is accessing the information, for what purpose and under what circumstances: when accessing EMR repositories for point-of-care uses, there may be consent and governance requirements that allow only named clinicians directly responsible for care to access details, whereas in the case of research queries an ethics committee may only allow access to data about deceased patients where the data has been anonymised or partly anonymised (pseudonymised).

Archetypes allow for the modelling of knowledge about discrete classes of data and may also provide a suitable foundation for applying security controls. There is a need to capture the policy specifications for the protection of those discrete data items, and it is proposed that this can be managed by the construction of a knowledge management framework. Part of the requirements analysis has led to the initial design of a new formalism, based on the theory that reusing the information model and Archetype approach will meet the protection requirements and policy based controls. This formalism is known as the *Secutype*. The function of the Secutype will be to bind policy-based control information to Archetypes, and allow that information to be used to assert those controls to software tools, some of which will require construction for the purpose (a de-identification component, for example). Figure 2 illustrates where Secutypes operate in the management of an EMR. On the next page, Figure 3 shows the structure of a Date of Birth Archetype with the encapsulated data value for the date of birth; Figure 4 shows a depersonalisation Secutype with the controlled data values that will be released, based on examples of policy that exist in the CLEF project.

An overhaul of existing editing tools for Archetypes will be required to support the new Secutype capabilities, and this is being conducted as part of a doctoral work by the first author. The results of analysing the working patterns of the informatics community

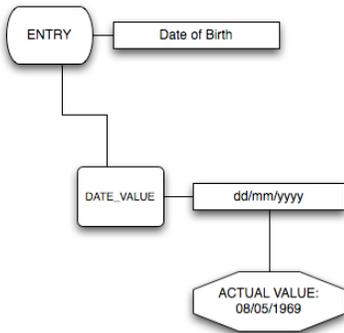


Figure 3. Date of Birth Archetype

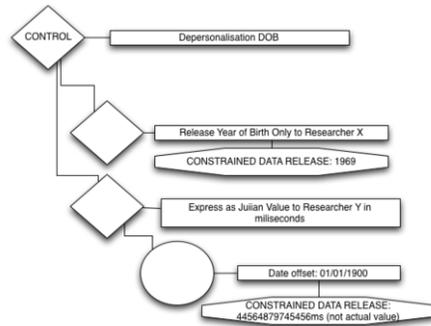


Figure 4. Depersonalisation Secutype

are significant: international collaboration on building information models and Archetypes appears to be popular and essential based on the work of communities such as the *openEHR* Foundation (www.openehr.org) where commentary by interested parties on Archetypes occurs via email. Another community developing models for the United Kingdom National Health Service has adopted a Wiki approach to collaboration (see <http://www.ehr.chime.ucl.ac.uk/display/nhsmodels/Home>), where any party (even outside the editing team) is permitted to leave comments in support or criticism of the current specification. However the Wiki used is a standard one that does not provide editing support for Secutypes in particular. The planned new Secutype editor uses a highly generic model for the capture of definition statements, and a Web application that provides a more focussed collaborative, editing environment is being developed for this. The model and application is being created based on the protection requirements and the needs of the informatics community already outlined.

After the model and application are completed, it is anticipated that there will be issues of practicality when they are applied to EMR systems, particularly when scaled to thousands of patients' records. These issues and possible solutions will be investigated as another part of the doctoral work, where Secutypes will be added to an advanced implementation of the EN / ISO 13606 record standard. Comparisons of system performance will be run with the Secutype components enabled and disabled across increasing numbers of patient records to discover the impact that Secutypes will have: once the practical issues have been established, potential solutions can be proposed, implemented and evaluated. Additional future research will explore how large, diverse and granular a set of Secutypes is needed within one healthcare organization or region.

3. Conclusions

This paper has proposed that a knowledge management approach is reasonable for the assertion of the required policy-based controls, and that a new formalism called Secutypes, based on the design principles of information models and Archetypes, can be used to capture the required details of policies and facilitate control assertions where needed. The paper has also identified anticipated issues of scalability and performance

once the Secutypes are applied to a running EMR, and indicated the proposals of further doctoral work on this subject.

Further research will provide a means to share details about security requirements for EMRs in the different use environments in which they can operate. A contribution of this work is anticipated to be the foundation of a library of Secutypes that might in the future facilitate consistent good practice across medical repositories within a nation or health system in which most policies ought to be the same. This will also support the widespread reuse of controls that already exist for managing access control, audit and data integrity, which currently require manual configuration based on a human-readable policy. It will allow for modularisation of the security in existing EMR servers so that policy can be automatically applied to fine-grained discrete data items, a requirement of modern data protection controls. Finally, there will be investigation into the performance impact of applying this control mechanism, and whether it is scaleable to the millions of patients' EMRs in the distributed computing environments used for large scale, national information technology projects.

References

- [1] D. Kalra, J. Milan, T. Austin, D. Ingram, D. Lloyd, D. Patterson, J. Grimson, W. Grimson, Synapses in Use: Supporting Cardiac Care at the Whittington Hospital. *Towards an Electronic Health Record Europe '98*. Medical Records Institute for the Centre for Advancement of Electronic Records Ltd. (1998), 306-312.
- [2] *openEHR Usage*: The *openEHR* Foundation. <http://www.openehr.org/shared-resources/usage/governments.html> (last accessed 15th November 2007).
- [3] *Health Level 7 Record Information Models*: Health Level Seven. www.hl7.org (last accessed 15th November 2007)
- [4] *openEHR release 1.0.1 Specification Road Map*: www.openehr.org/svn/specification/TAGS/Release-1.0.1/publishing/roadmap.html (last accessed 16th November 2007).
- [5] S. Garde, E. Hovenga, J. Buck, P. Knaup. Expressing Clinical Data Sets with openEHR Archetypes: A Solid Basis for Ubiquitous Computing. *International Journal of Medical Informatics* **76** (2007), 334-341.
- [6] *New Operational Procedures for NHS RECs Guidance for applicants to Research Ethics Committees*: Central Office for Research Ethics Committees (COREC), www.sehd.scot.nhs.uk/cso/AboutCSO/Ethics/COREC%20flyer.pdf 1st March 2004. (last accessed 15th November 2007).
- [7] *Information Governance – Connecting for Health*: www.connectingforhealth.nhs.uk/systemsand services/infogov (last accessed 16th November 2007).
- [8] M.Y.Becker, Information Governance in NHS's NPfIT: A Case for Policy Specification. *International Journal of Medical Informatics* **76** (5-6) (2006), 432-437.
- [9] T. Austin, The Development and Comparative Evaluation of Middleware and Database Architectures for the Implementation of an Electronic Healthcare Record. www.ehr.chime.ucl.ac.uk/download/attachments/71/Austin%2C+Tony+%28PhD+2004%29.pdf?version=1 (last accessed 16th November 2007).
- [10] D. Kalra, Clinical Foundations and Information Architecture for the Implementation of a Federated Health Record Service: www.ehr.chime.ucl.ac.uk/download/attachments/71/Kalra%2C+Dipak+%28PhD+2002%29.pdf?version=1 (last accessed 16th November 2007).
- [11] R. Anderson, Under Threat: Patient Confidentiality and NHS Computing: <http://www.cl.cam.ac.uk/~rja14/Papers/drugsandalcohol.pdf> (last accessed 16th November 2007).
- [12] M. Sloman, E. Lupu. Security and Management Policy Specification. *IEEE Network* **16** (2002), 10–19.
- [13] D. Kalra, P. Singleton, D.Ingram, J.Milan, J. MacKay, D. Detmer, A. Rector, Security and Confidentiality Approach for the Clinical eScience Framework (CLEF). *Methods of Information in Medicine* **44** (2) (2005), 193-197.
- [14] C. Haynes, G. Cook, M. Jones, Legal and Ethical Considerations in Processing Patient-Identifiable data without Patient Consent: Lessons Learnt from Developing a Disease Register. *Journal of Medical Ethics* **33** (2007), 302-307.

Improving Patients Privacy with Pseudonymization

Thomas NEUBAUER and Bernhard RIEDL
Secure Business Austria, Vienna, Austria

Abstract. e-Health requires the sharing of patient related data when and where necessary. Electronic health records promise to improve communication between health care providers, thus leading to better quality of patients' treatment and reduced costs. As highly sensitive patient information provides a promising goal (e.g., for attackers), there is an increasing social and political pressure to guarantee patients privacy. This paper presents the new system PIPE (Pseudonymization of Information for Privacy in e-Health), that differs from existing approaches in its ability to securely integrate primary and secondary usage of health data.

Keywords. Privacy, e-health, Security, EPR-CPR-EMR, Smart Cards

Introduction

The discussion of privacy is one of the fundamental issues in health care today and a trade-off between (i) the patient's requirement for privacy as well as (ii) the society's need for improving efficiency and reducing costs of the health care system. Electronic health records (EHR) were introduced over the past several years as a method for improving communication between health care providers and access to data and documentation, leading to better clinical and service quality [10]. The EHR promises massive savings by digitizing diagnostic tests and images. A study by the non-profit research organization Rand Corporation found out that adopting the EHR could result in more than \$81 billion in annual savings in the US, if 90% of the health care providers used it [5]. However, highly sensitive data is stored and handled in nationwide or european-wide medical systems that are often available over the Internet and hardly protected. As a result of the high sensitivity of medical data and due to an almost endless list of security breaches revealing patients' data (cf. [2]), there is an increasing social, legal and political pressure to prevent the misuse of health data.

It is the patients right to demand privacy, because the disclosure of medical data may cause serious problems for the patient. Insurance companies or employers could use this information to deny health coverage or employment. Therefore, legal acts demand the protection of health data (HIPAA [22], Directive 95/46/EC [6], Article 29 Working Party [7], Article 8 [3] of the European Convention for the Protection of Human Rights and Fundamental Freedoms, domestic acts, e.g., the Austrian Data Protection Act [15]). This paper evaluates existing privacy enhancing techniques (PETs) used for securing medical data and presents a novel approach for granting privacy through encrypted pseudonyms and authorization via encryption named PIPE

(Pseudonymization of Information for Privacy in e-Health). This new system differs from existing approaches in its ability to securely integrate primary and secondary usage of health data. By showing new concepts for data sharing, authorization and data recovery in case the user loses her access key, it provides a solution to security shortcomings of existing approaches.

1. Background

The increasing fear of data abuse as well as the need for compliance to the legal demands lead to the development of a variety of techniques for protecting patients' identity and privacy. In order to protect patients' privacy when using and transferring medical records a variety of approaches is currently used (cf. Figure 1 for a comparison of existing approaches regarding secondary use):

- Anonymization (cf. [14,12,21]) is the removal of the identifier from the medical data. It is realized by deleting the patient's identification data and leaving the anamnesis data for secondary use. As this approach does not establish a link between the anonymized data and its associated individual, it is the most secure way for granting privacy. Although this approach is often used in research projects due to its simplicity, it has the major drawback that primary use cannot directly profit from the results made in the research project (e.g., patients cannot be informed about actual findings such as newly developed medical treatment or major changes in the healing progress).
- Depersonalization (cf. [14]) is a technique similar to anonymization and comprises the removal of as much information as needed to conceal the patients' identity.
- Encryption (cf. [17]) assures patients' privacy by encrypting the anamnesis data with the patients' private key. As medical data tends to be very large (e.g., the image size of a x-ray is 6 MB, for a mammogram 24 MB or for a computer tomography scan counts up to hundreds of MB) encryption is a highly time-consuming operation. Beside that, encrypted data cannot be used for research projects (secondary use) without explicit allowance by the patient who has to decrypt the data and, thus, discloses his identity.
- Role-based access control (RBAC) (cf. [16]) can be used for restricting the data access to authorized persons. However, role-based access control models can be by-passed or compromised and persons with administrative roles still have un-restricted access to the whole database. Therefore, this technique does not provide enough security for protecting sensitive health data, because attackers - no matter if working inside the system or getting access from outside the system - could get full access to the database including the relations between patient's identifiers and their medical data.
- Pseudonymization (cf. [11,20,4,8,13,12,19]) allows an association with a patient only under specified and controlled circumstances. It is a technique where identification data is transformed into a and afterwards replaced by a specifier which cannot be associated with the identification data without knowing a certain secret. Existing approaches have shortcomings such as the concealment of the applied algorithms or the use of centralized patient-pseudonyms lists. Relying on a list is not secure, because an attacker, who gains access to this list, could establish an unauthorized relation between the

patient's identifier and his medical data. Existing approaches also neglect to provide efficient and secure fall-back mechanisms for recovering the key in order to re-establish access to the data if the patient lost his smart card and often depend on a single pseudonym.

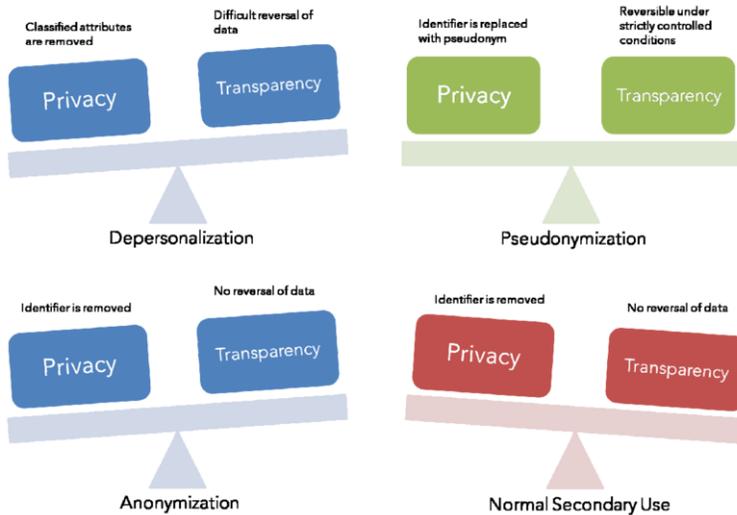


Figure 1. Comparison of existing approaches

2. Pseudonymization of Information for Privacy in e-Health

The goal of our architecture is to gain the optimal trade-off between security on the one hand and usability and performance on the other hand. PIPE comprises the following components: The Storage is divided into two separate storage systems (e.g., databases), where one is related to identification data and the other one is related to pseudonymized data as well as the associated pseudonym *PSN*. The central logic (e.g., a server) provides an interface between the central storage and the clients for the purpose of saving and loading the data. Note, that all private keys are identified as e (e.g., the patients' inner private key will be named $e'A$) and all public keys are denoted with d . The patient (A) has full access to his data via the central logic by using his security token (e.g., smartcard with a PIN). The patient may provide relatives (B) with his inner private key, which will then be encrypted with the relative's symmetric key. By doing this, the relative gets access to all data of the patient, until the patients' inner private key is changed. A health care provider (C) can be authorized by the patient to access a subset of his anamnesis datasets. The health care provider may also share one or more entries in the pseudonymized database with the patient. The research lab has only access to the concealed data CD on the server system via the central logic for the purpose of analysis needed for improving the efficiency of clinical trails, the medical treatment, or medication. The operator(-team) (O) holds secrets on behalf of the system. This role assures that if a patient loses or destroys his smartcard, the access to the system can be restored by a team of operators.

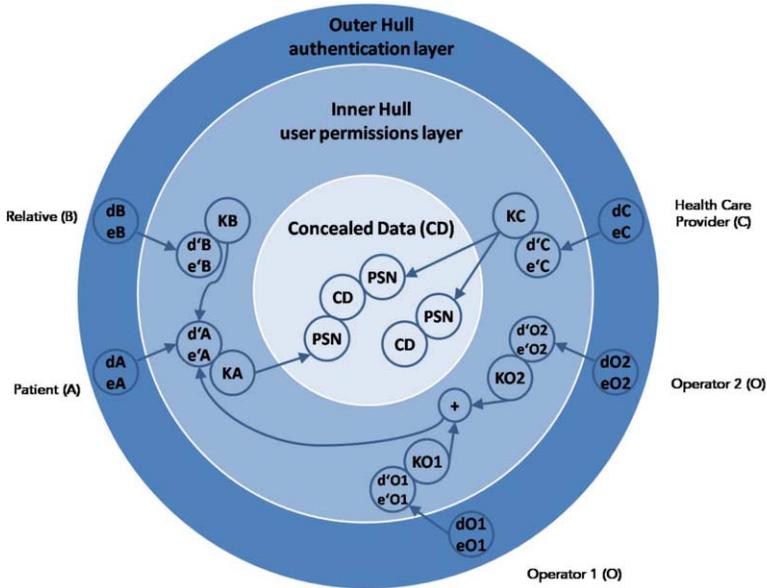


Figure 2. The Hull Architecture of PIPE

Figure 2 shows the hull architecture of PIPE. In the most outer layer (user permissions layer) every user possesses a security token (e.g., in our prototype we used smart cards as security tokens) to access the secrets of the next inner hull. The anamnesis data is stored pseudonymized in the most inner hull (concealed data layer).

Any medical dataset is associated with one or more unique pseudonyms. As the patient is the owner of the data, she is the only person who holds the so-called root-pseudonym PSN_0 . All other pseudonyms PSN_j are disjunct for any patient, health-care-provider and anamnesis combination. If, for example, two health care providers have been authorized to access a specific anamnesis CD , three pseudonyms (PSN_0 , PSN_1 and PSN_2) exist. All of these pseudonyms are stored encrypted with the particular users' inner symmetric keys, whereas the plain-text medical data is associated with the plain-text pseudonyms. Thus, the pseudonymization can be reversed by using the patient's inner symmetric K_A or any authorized health care provider's C inner symmetric key K_C . To get access to these particular keys, the authorized users' inner private key has to be used. If a patient A shares her secret of the inner hull, she consequently provides access to all her data, if not additionally revised by an access control model. We define two main roles which may hold an encrypted copy of the patient's inner hull secret, her inner private key $e'A$. Firstly, a relative B may encrypt the patient's inner private key with her inner public key $d'B$. Thus, she is also able to decrypt the patient's inner symmetric key, until the patient would change it.

Secondly, as a user's smart card may be lost, destroyed, stolen, compromised or just worn-out, there is the need to keep a backup of the user's inner private key, because otherwise the user's data would not be accessible any more. This backup keystore has to be secured and protected against fraud. In our prototype we applied Shamir's threshold scheme for securely sharing secrets [18] between a set of operators O , who are randomly assigned to hold a part of the users' secrets. Since users need a fall-back mechanism in case they have lost their smart card, operators O hold the users'

inner private keys on behalf of the patient. To avoid misuse of these rights, we applied a two-step variant of Shamir's threshold scheme [18]. Following Shamir, two parameters can be defined for sharing a secret, (i) the number of shares n and (ii) the amount of shares k , that are necessary to re-establish the certain secret. The higher the number of issued shares compared to the number of shares in total, the higher the security, assuming the operators are randomly assigned, each holding one share. Of course, decrypting operations conducted by humans cause higher costs than performed by machines. To decrease the costs for establishing a backup keystore, we propose a combination of human operators and machine operators, for example Hardware Security Modules. Secure authorizing is another vital function of our system to assure that the users are in full control of their actions at any time. If there is for example the need to authorize an additional user for an anamnesis, another pseudonym has to be linked with the certain anamnesis, which the data owner and the additional user hold together.

3. Conclusions

Health care providers require the sharing of patient related data (e.g., using EHRs) in order to provide efficient patients treatment. The implementation of EHRs does not only promises a higher level of service quality for the patients, but also reduces costs for social insurance systems and therefore for the society. As highly sensitive and personal information is stored and shared within these highly interconnected systems, there is increasing political, legal and social pressure to guarantee patients' privacy. This paper gave an overview of the shortcomings of existing approaches, such as their dependence on a centralized patient-pseudonyms list, a life-long pseudonym or the concealment of the used algorithm. Based on these shortcomings this paper presented a secure architecture for the combined primary and secondary usage of health-related data. Relying on the encryption of the systems relations, PIPE assures that the patients are in full control of their data. The approach provides research organization with data for improving medical treatment or clinical paths. The proposed fall-back mechanism allows recovering the patient's secret key in case he lost his smart card. Workflows that demand the authorization of one of the systems' participants demand a the use of a secure viewer (cf. [9,1]) in order to guarantee confidentiality, integrity and non-repudiation (e.g., that only the intended person is authorized). The information about the person that should be authorized must be presented to the user without any manipulations (e.g., committed by a man-in-the-middle). Therefore, further research will extend our concept by integrating a secure viewer mechanism and present the results of some case studies. The case studies will focus on the integration of our concept into existing workflows and systems in the health sector in compliance with legal requirements.

Acknowledgments

This work was performed at Secure Business Austria, a competence center that is funded by the Austrian Federal Ministry of Economics and Labor (BMWA) as well as by the provincial government of Vienna.

References

- [1] Adil Alsaid and Chris J. Mitchell. Dynamic content attacks on digital signatures. *Information Management & Computer Security*, 13(4):328–336, 2005.
- [2] Privacy Rights Clearinghouse. A chronology of data breaches.
- [3] Council of Europe. *European Convention on Human Rights*. Martinus Nijhoff Publishers, 1987.
- [4] G.J. de Moor, B. Claerhout, and F. de Meyer. Privacy enhancing technologies: the key to secure communication and management of clinical and genomic data. *Methods of information in medicine*, 42:148–153, 2003.
- [5] Frank R. Ernst and Amy J. Grizzle. Drug-related morbidity and mortality: Updating the cost-of-illness model. Technical report, University of Arizona, 2001.
- [6] European Union. Directive 95/46/ec of the european parliament and of the council of 24 october 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal of the European Communities*, L 281:31–50, 1995.
- [7] European Union, Article 29 Working Party. Working document on the processing of personal data relating to health in electronic health records (EHR), February 2007.
- [8] J.R. Gulcher, K. Kristjansson, H. Gudbjartsson, K., and Stefansson. Protection of privacy by third-party encryption in genetic research. *European journal of human genetics*, 8:739–742, 2000.
- [9] Hanno Langweg. Malware attacks on electronic signatures revisited. In J. Dittmann, *Konferenzband der 3. Jahrestagung Fachbereich Sicherheit der Gesellschaft für Informatik*, pp. 244–255, 2006.
- [10] S. Maerkle, K. Koechy, R. Tschirley, and H. U. Lemke. The PREPaRe system – Patient Oriented Access to the Personal Electronic Medical Record. In *Proceedings of Computer Assisted Radiology and Surgery, Netherlands*, pages 849–854, 2001.
- [11] Robert L. Peterson. Patent: Encryption system for allowing immediate universal access to medical records while maintaining complete patient control over privacy. *US Patent US 2003/0074564 A1*, 2003.
- [12] Andreas Pfitzmann and Marit Koehntopp. Anonymity, Unlinkability, Unobservability, Pseudonymity, and Identity Management -A Consolidated Proposal for Terminology. In *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2005.
- [13] Klaus Pommerening and Michael Reng. *Medical And Care Compunetics 1*, chapter Secondary use of the Electronic Health Record via pseudonymisation, pages 441–446. IOS Press, 2004.
- [14] Alan Rector, Jeremy Rogers, Adel Taweel, David Ingram, Dipak Kalra, Jo Milan, Peter Singleton, Robert Gaizauskas, Mark Hepple, Donia Scott, and Richard Power. Clef -joining up healthcare with clinical and post-genomic research. In *Proceedings of UK e-Science All Hands Meeting*, pages 203–211, 2003.
- [15] Republic of Austria. Datenschutzgesetz 2000 (DSG 2000), BGBl. I Nr. 165/1999, 1999.
- [16] R.S. Sandhu, E.J. Coyne, and C.E. Feinstein, H.L. and Youman. Role-based access control models. *Computer*, 29(2):38–47, 1996.
- [17] Volker Schmidt, Werner Striebel, Heinz Prihoda, Michael Becker, and Gregor De Lijzer. Patent: Verfahren zum be-oder verarbeiten von daten. *German Patent, DE 199 25 910 A1*, 2001.
- [18] Adi Shamir. How to share a secret. *Commun. ACM*, 22(11):612–613, 1979.
- [19] Ka Taipale. Technology, Security and Privacy: The Fear of Frankenstein, the Mythology of Privacy and the Lessons of King Ludd. *International Journal of Communications Law & Policy*, 9, 2004.
- [20] Christian Thielscher, Martin Gottfried, Simon Umbreit, Frank Boegner, Jochen Haack, and Nikolai Schroeders. Patent: Data processing system for patient data. *Int. Patent, WO 03/034294 A2*, 2005.
- [21] Denise Thomson, Lana Bzdel, Karen Golden-Biddle, Trish Reay, and Carole A. Estabrooks. Central Questions of Anonymization: A Case Study of Secondary Use of Qualitative Data. *Forum Qualitative Social Research*, 6:29, 2005.
- [22] United States Department of Health & Human Service. Hipaa administrative simplification: Enforcement; final rule. *Federal Register / Rules and Regulations*, Vol. 71, No. 32, 2006.

Mobile Health Requires Mobile Security: Challenges, Solutions, and Standardization

Peter PHAROW¹ and Bernd BLOBEL

eHealth Competence Center, Regensburg University Hospital, Regensburg, Germany

Abstract. Extended communication and advanced cooperation in a permanently growing healthcare and welfare domain require a well-defined set of security services provided by an interoperable security infrastructure based on international and European standards. Any communication and collaboration procedure requires a purpose. But such legal purpose-binding is definitely not the only aspect to carefully be observed and investigated. More and more, aspects of security, safety, privacy, ethics, and quality reach importance while discussing about future-proof health information systems and health networks – regardless whether local, regional or even pan-European networks. During the course of the current paradigm change from an organization-centered to a process-related and to a person-centered health system, different new technologies including mobile solutions need to be applied in order to meet challenges arising from both legal and technical circumstances. Beside the typical Information and Communication Technology systems and applications, the extended use of modern technologies includes large medical devices like, e.g., MRI and CT but also small devices like sensors worn by a person or included in clothing. Security and safety are on top of the priority list. The paper addresses the identification of some specific aspects like mobile technology and safety when moving both IT and people towards mobile health aiming at increasing citizens and patients awareness, confidence, and acceptance in future mobile care - a world often still beyond the horizon.

Keywords. eHealth, mHealth, Personal health, security, safety, mobility.

Introduction

Advanced information and communication technology (ICT) solutions are an important pre-requisite for establishing today's integrated care solutions. Disease management programs (DMP), public health and surveillance systems, and personal care settings are other aspects to be addressed in this context. The ICT infrastructure is hereby a crucial part of the organizational infrastructure as such. While the healthcare and welfare system undergoes the paradigm change from shared care to personal care, the infrastructure needs to undergo a similar migration. It's not 1970s monolithic dinosaur systems, it's not 1980s decentralized systems, and it's not PC-based application environment that may help responding to new challenges. It's the new world of mobile, and mobilizing, sensors, devices, services, systems, patients and health professionals. The paper aims at addressing mobile health technology security and safety problems.

¹ Corresponding Author: Peter Pharow, eHealth Competence Center Regensburg, Regensburg University Hospital, Franz-Josef-Strauss-Allee 11, D-93053 Regensburg. Phone: +49-941-944-6767. Fax: +49-941-944-6766. Email: peter.pharow@klinik.uni-regensburg.de. URL: <http://www.ehealth-cc.de>

1. Current Healthcare Requirements

The healthcare and welfare systems in all countries around the globe are faced with similar problems. Citizens and patients express increasing expectations while available resources from health funds decrease permanently. In order to cover this funds shortage, patients are obliged to pay more and more services from their own pocket. For an advanced cooperation and collaboration, health networks are established crossing not only borders of domains but increasingly also national boundaries. The society experiences a paradigm change from “Shared Care” towards “Personal Care” requiring advanced communication infrastructures [1]. The care process itself is typically organization-centered meaning that all treatment processes are organized by exactly the hospital or GP the patient has visited. Information exchange is rather rare in this context. Some countries have already started developing a concept of process-oriented care. It’s not a single organization; it’s the disease itself that is in the center of processes including information storage and exchange. All health professionals involved in patient’s care regardless of their affiliation have access to, and provide, case-related medical data items. To meet the comprehensive paradigm of Personal Care, care provision needs to be further developed towards a patient-centered (or rather person-centered) approach moving the citizen into the center of all care processes [2].

This new and enhanced approach needs to invite and incorporate not only health professionals and patients (citizens) but also many new stakeholder groups like, e.g., health politicians, standardization experts, large as well as medium and small (SME) enterprises, professional and citizen associations, chambers, and other organizations. In this respect, eHealth is considered advanced and extended health telematics based on new telemedical technology requiring an extended communication and collaboration. The personal(ized) health (pHealth) provision itself obviously addresses the citizen (patient, client, consumer) being moved to the center of all processes requiring an extended personal and personalized communication among all principals. Mobile health (mHealth) fosters processes to be performed in mobile environment meaning that the network beside traditional applications and systems increasingly contains mobile components (wearable sensors, RFID tags, micro-technology as well as self-organizing components and systems). Communication in mHealth is mainly performed in a mobile network environment [3].

2. Materials and Methods

So what does mobile mean in this respect? What does mobility mean? Is there a sufficient level of awareness, confidence and acceptance among all principals involved for the new levels of health-related security and safety requirements in a speedy world that gets more and more mobile? Is there a technology in place that is promising enough to fulfill or at least strongly support the advanced mHealth requirements?

2.1. *Mobile Technology and Mobile Devices*

In the virtual world of mHealth, communication and collaboration partners typically do not know each other personally. Increasingly, such communication is performed via protocols for the mobile world (e.g. UMTS). One of the basic requirements is that of a secure, trustworthy, reliable communication using insecure connection means and

media like, e.g., the open and policy-free Internet. Trustworthiness is mainly based on technical mechanisms and algorithms providing communication security services such as identification, authentication, integrity, non-repudiation of origin (authenticity) and receipt, availability and confidentiality as well as application security services such as confidentiality, privacy based on privilege management, authorization, access control, anonymization, pseudonymization but also integrity, accuracy, accountability, auditability and traceability. Identity management (IDM) is a crucial service for most of the aforementioned ones. It includes any actors (persons, organizations, systems, devices, etc.). Other infrastructural services needed are policy and role management.

Devices are among the top priorities in advanced mobile environments. There's almost no application area in healthcare and welfare without having applied appropriate mobile devices technology. From emergency to treatment to rehabilitation to home-care, personal(ized) medical devices consistently help, among others, treating and monitoring patients at risk or patients at home (see Figure 1).



Figure 1. Medical mobile devices apply in almost all treatment phases

Quite often especially in the mobile domain, proprietary solutions, outdated approaches, and traditional principles are still in routine use for communicating sensitive information. The awareness level of principals is not significantly high, so real and potential threats and risks are often not addressed. The respective security and risk management is therefore quite weak. Standards like ISO/IEC 17799, ISO/IEC 27799 or the ISO 27000 family are often not known even by professionals [4], [5].

2.2. Patient and Citizen Requirements

The aspect of an active citizen and patient involvement and patient empowerment in the healthcare and welfare processes is an important pre-requisite for achieving the success expected by many developed countries. Health cards and other trustworthy and secure (personalized) devices do play an important role in this process. They allow developing integrated eHealth environments involving health informatics and patient integration. Keywords to foster these processes are health card, secure token and device-related technology but also the role of these tokens and devices in the process of patient integration, their security functionality for data, process and actors' integrity, the problems concerning access rights, and related data protection and privacy aspects.

From the aforementioned point of view, healthcare systems will increasingly profit from engaged citizens; and engaged patients will have better outcomes. New health

policies, increasing patient needs and decreasing health budgets necessitate patient empowerment. Patient involvement has to take place at different levels within the healthcare process -for example- at the level of care but also at the political level. At the level of care, the interest in citizen engagement is presently changing from past practices -where patient involvement meant that the patient had to go along with the physician's advice- to a collaborative partnership -where the citizen is considered a real partner within the medical workflow. Also at the political level patient group representatives are increasingly considered valuable partners [6], [7].

2.3. Applied Security Requirements

Secure and reliable communication in a mobile world has applied security requirements coming from different viewpoints. The first one is the "traditional" security domain. Mobile communications address both communication and application security services. But getting mobile definitely brings some important additional requirements addressing, i.e., also electrical, mechanical, and personal safety aspects.

Parameters like, e.g., robustness and self-adaptability is a major key for usability and usefulness in mobile terms, as it is power consumption as well as availability and distance of communication. Mobile devices need to be simple, efficient, handy, and scalable. They need to be safe from electrical, electronic, and personal standpoint. Medical mobile devices need to address these issues, and some other requirements on an even higher quality and reliability level.



Figure 2. Mobile Device for Iris Scan (Biometrics)



Figure 3. Mobile device for Fingerprint Scan

2.4. Technical Security Approaches

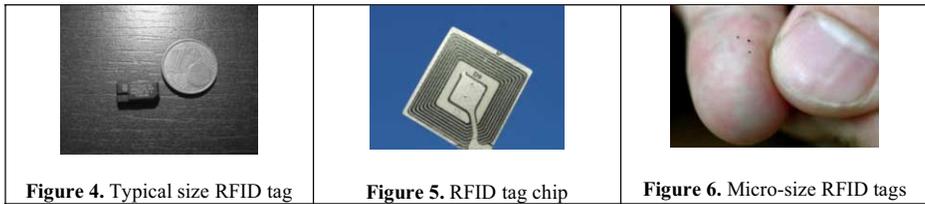
Mobile communication requires the same -or at least a similar- level of security as ordinary communication using typical infrastructures like LAN or MAN. Thus, mobile security solutions can be based on well-known and well-established symmetric (DES, 3DES, IDEA, AES) and asymmetric (RSA) cryptographic algorithms as well as key traditional exchange protocols, e.g., following the Diffie-Hellman approach [8]. On the other hand, the specific security aspects of a reliable wireless communication require specific modifications in the related protocols, mechanisms, and algorithms. Among others, data to be transferred are typically pre-processed and comprised in order to save time, bandwidth, and power consumption. Micro-technology is therefore used as one of the preferred technologies for advanced security and safety solutions in the context of eHealth, pHealth and especially mHealth.

3. RFID and NFC as a new Chance and Challenge

Radio Frequency Identification (RFID) and Near Field Communication (NFC) are technical solutions to be applied in healthcare and welfare. Information systems combine ordinary information services with the use of RFID for identifying either goods (drugs, tools, tissues, samples) or even persons (health professionals, patients). Data is exchanged between passive RFID tag and readers in a few centimeters range.

From the health safety point of view, RFID is considered to possibly cause certain health problems. The ubiquitous presence of electromagnetic fields (EMF) is often expected to cause effects of electric or thermic nature to biological systems. In reality, RFID-related EMF values have a considerably low influence on humans. Side effect, however, can appear and shall specifically be considered when pace makers, brain stimulators or other implants are applied. In general, health and personal safety are not affected by RFID and NFC. Nevertheless, technical solutions need specific healthcare and welfare profiles to comprehensively address the health-related safety aspects.

But there's indeed a problem with RFID as such. Micro-technology is able to deliver RFID tags that are small enough to be provided in liquid even without be noticed by the persons. This is an application area that combines safety aspects and privacy aspects because unexpected and unwanted profiling (tracking) and monitoring will be possible which affects the privacy rights of the persons.



Applying technical standards from other domains in eHealth and mHealth is a challenging task. The US NIST standard, e.g., aims at describing, among others, the applied use of RFID technology in a healthcare environment [9]. Different application fields are identified. They cover asset management, tracking, authenticity, matching, process control, access control, and payment. So many different domains -applying advanced RFID technology- work together forming a supply chain.

4. Discussion of Results and Strategies

Mobile communication often suffers from limited bandwidth. Algorithms pre-process and comprise recorded data. Proprietary protocols do have different serious weaknesses. New protocols and services like, e.g., WPA (Wi-Fi Protected Access), WPA2, and Bluetooth still base on the family of IEEE 802.1x standards but address some of the weak points of the standards themselves and try to overcome the most important drawbacks and pitfalls. Adding a real authentication service (RADIUS) allows for achieving a security level that is (almost) the same compared with ordinary LAN security solutions. The old-fashioned WEP standard is out, WPA und WPA2 address the minimum requirements for a secure mobile communication. These tendencies in the mobile world have significantly influenced the further development of the IEEE

802.11i standard. It's still a long way to go; the fast application of wireless technology (PDA, UMTS, biomedical devices) has by far not reached its final level.

5. Conclusions

Enhancing mobility has a manifold meaning. It's not just the use of mobile phones for health-related purposes, and it's not just the mobility of both patients and health professionals. A mobile environment contains self-organizing systems and components along with mobile devices, tools, sensors, and much more. Patient empowerment means they have to take on their share of responsibility. Citizens have to realize that healthcare is a concern of theirs even while they are not in need of it. This means they have to understand the system. Awareness, confidence, and acceptance range high in future care settings. Standardized and standards-based mobile technology needs to support these processes. Security and safety supports application of new technologies.

The paper addresses awareness of security, safety and quality requirements and solutions in ubiquitous health settings based on the pHealth paradigm. Basic services such as ID management and privacy aspects have been discussed, while services such as policy bridging a still under development.

Acknowledgement

The authors are in debt to the European Commission for supporting and funding the "BioHealth" project as well as other partners and organizations (ISO/TC 215, CEN/TC 251, EFMI WGs CARDS and SSE) for their permanent support and cooperation.

References

- [1] Blobel B, Pharow P (Eds.): *Advanced Health Telematics and Telemedicine*. The Magdeburg Expert Summit Textbook. Series "Studies in Health Technology and Informatics" Vol. 96. IOS Press, Amsterdam 2003
- [2] Blobel B: *Analysis, Design and Implementation of Secure and Interoperable Distributed Health Information Systems*. Series "Studies in Health Technology and Informatics" Vol. 89. IOS Press, Amsterdam 2002.
- [3] Pharow P, Blobel B: *Von eHealth zu mHealth – die Sicherheitsanforderungen in der mobilen Welt*. Proceedings of GMDS 2007. <http://www.med-ges-2007.de/> (in German).
- [4] ISO/IEC 17799:2005 "Information technology -- Security techniques -- Code of practice for information security management". Now also available as ISO/IEC 27002:2005 "Information technology -- Security techniques -- Code of practice for information security management".
- [5] ISO/DIS 27799 "Health informatics -- Information security management in health using ISO/IEC 17799".
- [6] International Alliance of Patients' Organizations (IAPO). <http://www.patientsorganizations.org> (last accessed February 5th, 2008)
- [7] IAPO Policy Statement on Patient Involvement. <http://www.patientsorganizations.org/showarticle.pl?id=590&n=962> (last accessed February 5th, 2008)
- [8] Pharow P, Blobel B: *Security Infrastructure Requirements for Electronic Health Cards Communication*. In: Engelbrecht R., Geissbuhler A. (Eds.): *Connecting Medical Informatics and Bio-Informatics*. Proceedings of MIE 2005. Series "Studies in Health Technology and Informatics" Vol. 116. IOS Press, Amsterdam 2005.
- [9] US National Institute for Standards and Technology (NIST): *NIST Special Publication 800-98: Guidelines for Securing Radio Frequency Identification (RFID) Systems*. http://csrc.nist.gov/publications/nistpubs/800-98/SP800-98_RFID-2007.pdf (last acc. Feb 5th, 2008).

Towards Dynamic Access Control for Healthcare Information Systems

Lillian RØSTAD^a, Øystein NYTRØ^a

^aDepartment of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway

Abstract. Access control is a key feature of healthcare information systems to protect the privacy of patients and to ensure access to information as required by healthcare professionals. A problem with many existing access control mechanisms is their static nature. In this paper we propose combining workflow information from medical guidelines, observations and audit logs to create dynamic access rules that are adapted to the actual workings of a hospital. Our aim is to help minimize the use of “break the glass” access.

Keywords. Security, Data protection, Evidence based guidelines

Introduction

Access control is one of the key features of health care systems. Access control is about restricting as well as ensuring access to information. These are two inherently different viewpoints. For privacy it is important that access is only granted when there is a legitimate need. For availability it is equally (some would argue more) important that access is granted to all information required to provide the best possible care. The goal of the work presented here is to narrow the gap between these viewpoints, by proposing a method for dynamic access control rules that adheres to the actual flow of work and responsibilities in a hospital setting.

1. Access Control Concepts

Access control is about enforcing rules on which operations a user is allowed to perform on a resource (eg. information) in a system. There are several different access control models. The most common ones are mandatory access control (MAC), discretionary access control (DAC), and role-based access control (RBAC) [1]. RBAC is the preferred model used in many implemented access control mechanisms in health care systems and serves as the foundation for the ideas presented here.

1.1. Role Based Access Control

The concept of Role Based Access Control [1] has gained increasing popularity over the last decade. In RBAC a set of roles is created that corresponds to job functions in an organisation. Each role consists of a set of access rules. Rather than assigning access rules directly to a user, a user is linked to the role and thus has all the access rules associated with that role. Several users may be assigned to the same role, and one user may take more than one role. A typical RBAC role in a health care system would be that of a nurse. The nurse role consists of access rules that correspond to the information access a nurse needs to perform his job.

The main advantages of RBAC are ease of administration, flexibility, and scalability. RBAC is considered a good fit when there are considerably fewer roles than employees in an organisation. When a new nurse is hired there is no need to create a specific access profile for him – he can simply be assigned to the existing nurse role. This scales well as it is easy to add more nurses as an organisation grows, and it is flexible because changing the access rules for all nurses only requires changing one role.

Some health care systems [2] combine a role with the user's place of work to make access decisions. In short, this means that a nurse only has access to users that are currently admitted to the ward where he works. Dynamic RBAC is extended to assign roles temporarily, according to work shifts or work processes.

1.2. Optimistic security

In [3] Povey proposes the concept of *optimistic security*. The key feature in an optimistic security mechanism is the use of *retrospective control*. There are no access rules that are enforced when a request is made. The concept relies on the ability of someone to examine the logs later and determine if the access was legitimate. Auditing and traceability therefore are keys to enforcing optimistic security. Povey argues that optimistic security is well suited for systems such as healthcare where there may be situations when a user needs to exceed his normal privileges.

Optimistic security exists in many healthcare systems as a “break the glass” mechanism intended to be used in emergency situations. A study [2] has looked into use of the “break the glass” mechanism in a system where normal access control is enforced as a combination of role and workplace as explained earlier. In the study audit data was collected for one month's use of the system at eight hospitals. The study found that 54% of the patients admitted in this time period had their record accessed using the “break the glass” mechanism. Out of all accesses made in this period, 17% were performed using the “break the glass” mechanism. These findings strongly suggest that the rather static approach to access rules (role and ward) does not perform very well in a dynamic hospital setting.

The 17% accesses resulted in almost 300 000 entries in the audit logs. The study also found that there were no automatic audit analysis tools in place. The amount of audit trails and the absence of tools make the task of analyzing audit trails for retrospective control impossible. A condition for optimistic security to work, is that the amount of use is minimal so manual review is realistic.

In health care there will always be situations where availability of information is crucial and “break the glass” mechanisms are needed. One example is emergency situations when there is no time to properly register the patient in the administrative

systems, which often is a requirement for normal access rules to apply. The goal is therefore not to completely eliminate the use of “break the glass”, but to reduce the use to an amount where it is feasible to perform retrospective control. One approach towards this goal is developing access control mechanisms that are better adapted to the actual workings of a hospital and are dynamic in the sense that they are able to change and adapt as situation and context change. We will explore this idea further in this paper.

1.3. Dynamic access control – related work

As stated earlier, a problem with many access control rules in health care is the “define once – use always” approach and the lack of dynamic properties and adaptability. Several extensions to RBAC have been proposed to include dynamic properties. Examples include *role delegation* [4] and *context-sensitivity* [5]. Role delegation allows a user to delegate her role to another user to transfer responsibilities either permanently or time-limited. In the proposed context-sensitive RBAC models, context is used to activate and deactivate roles. A user may have a large pool of assigned roles and only a subset of these may be activated at any given time. Context properties may be used to regulate the activation of a role. E.g info about work schedule may be used to activate roles depending on time and place of work.

Though some propositions have been made on how to make RBAC more dynamic, a discussion of exactly what properties or values may be used remains. In the remainder of this paper we propose combining established best practices (medical guidelines), collected observational data, and audit data to learn patterns of information used in healthcare and apply these patterns to create access control rules that will help minimise use of «break the glass» access.

2. Workflow knowledge

Medical guidelines, work plans, observed behaviour, and audit data all contain information about workflow in healthcare. While medical guidelines are the idealised version of the medical activities related to a problem, observational and audit data reflects what actually happens [6]. Moreover, guidelines do seldom assign roles or resources. However, by combining these sources of knowledge we can create a coherent view of enacted workflows in healthcare, with an emphasis on information access requirements that may be utilized for access control.

In this section we discuss medical guidelines, observation data and audit logs separately and provide motivational examples of how this information may be used for access control purposes.

2.1. Medical Guidelines

A medical guideline (MG) for a given diagnosis contains information about best practice course of treatment developed by experts in the field. Guidelines may exist both as an informal collection of information and in a more formalised, structured manner. MGs often include temporal and event information that implies information needs that may be utilized for access control purposes.

An example of a guideline for treatment and observation of Gestational Diabetes Mellitus (GDM – a form of diabetes found in pregnant women), encoded in the Asbru language for computer-interpretable medical guidelines, is available at [7]. Use of the guideline is initiated if a glucose tolerance test in the third trimester shows a blood sugar level between 140 and 200 mg/dl. The guideline consists of three main parts: glucose monitoring, nutrition, and insulin therapy.

The Asbru guideline for GDM contains both temporal and contextual information that may be used for access control:

- Periodic information needs: visits to physician while under treatment every 1-4 weeks (specific value set for a patient). The EPR does not need to be accessible to the physician in-between visits.
- Events that trigger information needs: when blood sugar readings are too high the patient needs to visit her physician and review treatment. The patient record should be made accessible to the physician when too high readings occur.

2.2. Observational data – empirical grounding of guidelines

Guidelines are constructed by experts and represent idealized treatment processes – what is expected to happen given a diagnosis. In reality, each patient and care process is unique; furthermore, a complex problem will require that different guidelines are combined. A guideline may serve as a starting point, but will often need to be adapted to the specific situation at hand. In [6] the authors discussed how to use methodical observations of clinical care situations to improve guideline implementation.

An observational study was carried out in the summer of 2005. Two medical students observed clinicians at work in the pre-rounds meeting and ward rounds. They took detailed notes of who were present, the subject of discussion (patient), information sources (written/electronic and oral), and specifics about what type of information was used. In each observation session they followed one clinician and from her viewpoint they noted who else were present and what role they had in the situation. We have reviewed these data to construct an example of how observational data may be used to create patterns of information/needs, shown in Figure 1.

Due to space limitations, Figure 1 shows only the first few interactions in the pre-visit meeting, but it is sufficient to serve our purpose as an illustrative example. In this case they are discussing patient NN. The patient is new to the doctor so the nurse fills him in on some background info. Several information sources are used – some are paper-based (the patient list and the patient chart) and some are computer-based information systems (the electronic patient record (EPR) and the radiology imaging system (IDS)). The figure illustrates communicative acts between the actors present and the actors and the information sources they use. Roles are used to label the actors. This figure illustrates how observation may be used to uncover information needs in specific situations with a specific diagnosis (in this case heart failure), and link these to roles. Though not shown in Figure 1, the observational data shows that the diagnosis changes as test are being done and test results received and reviewed, as is very common. Through observational studies we can examine these transitions and study transfer of responsibilities and access requirements related to this.

Pre-rounds meeting : patientNN

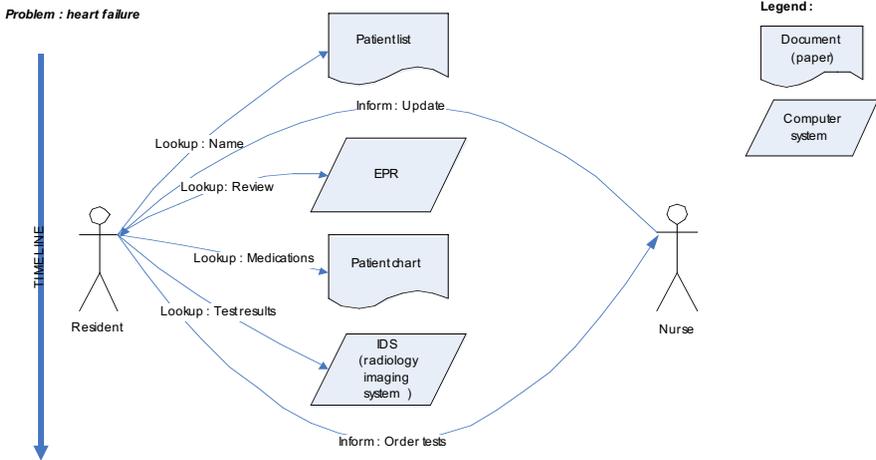


Figure 1 – Information needs in pre-rounds meeting

Even if observations provide real-world examples that may be collected over time, generalized, and used to improve guidelines, they still only give us a relatively high-level view. To complete this picture and get detailed and accurate information about information accessed and actions performed, we turn to the audit logs.

2.3. Usage patterns from audit logs

Most health care systems keep complete history; of changes in information and of user actions. The purpose is to always be able to roll back to a previous state, and to have complete traceability. This means that there exist audit logs with very detailed traces of user actions: the user's role at the time, what information was accessed, for which patient and what actions were performed [2]. From these audit logs it is possible to create generalized usage patterns per role. If a system allows “break the glass” access, it is also common to require the user to provide a reason for doing so and keep a log of these reasons as well [2]. We suggest utilizing this information for access control by:

1. Examine the reasons for using “break the glass” – any reasons that occur often should be considered as candidates for inclusion in the access control rule set.
2. Look for common usage patterns that describe workflows inwards. Examples include:

Temporal patterns

If action X occurs – then action Y occurs within Z time.

Responsibility patterns

If action X is performed by Role A – then action Y is performed by role B.

Location patterns

If action X is performed at ward 1 – then action Y is performed at ward 2.

Situation patterns

Role X is in situation S in a guideline, and requires specific information.

3. Discussion

“Break the glass” access is necessary to handle unexpected situations, but it constitutes a security risk and may be misused. The ideas presented here aim at minimizing the need for glass-breaking and making retrospective control feasible.

In access control, the main concern is privacy, where access should only be granted to the information required by an actor in any situation. Clinicians may well disagree with this from the viewpoint that it is better to have broad access. In this paper we therefore suggest an approach to access control that combines guidelines and learning from observations and logs. The goal is to take another step towards the goal of having access mechanisms that support the work of care providers, while protecting the privacy of patients.

The approach presented here is not another “do once – use forever approach”. It is fundamental to this idea that observing, learning, and improving should be a continuous process, allowing access rules to adapt to a dynamic, ever-changing environment.

4. Conclusion and future work

In any clinical situation, the information about a patient can be ordered along a continuum from highly relevant, via interesting, to irrelevant, and at the other extreme; illegal according to laws of privacy. Being able to sort correctly may mean life and death. The main problem facing today’s busy clinician is avoiding irrelevant information and at the same time getting access to relevant information. In this perspective, relevance ranking and access control depend on the same knowledge about situation, role, guideline, and care process. We believe that optimistic access control, based on analysis and learning from practice as intended and as enacted, is a first step towards both effective relevance ranking and optimal access control.

References

- [1]. Ferraiolo, D.F., D.R. Kuhn, and R. Chandramouli, *Role-Based Access Control*, 1 ed. Computer Security Series. 2003, Boston: Artech House Publishers, ISBN: 1580533701
- [2]. L. Røstad, O. Edsberg, *A Study of Access Control Requirements for Healthcare Systems Based on Audit Trails from Access Logs*, Proceedings of the 22nd Annual Computer Security Applications Conference, Miami, USA, 2006.
- [3]. D. Povey, *Optimistic security: a new access control paradigm*, Proceedings of the 1999 workshop on New Security Paradigms, Ontario, Canada, 1999.
- [4]. Zhang, L., G.-J. Ahn, and B.-T. Chu, *A role-based delegation framework for healthcare information systems*. Proceedings of the seventh ACM symposium on Access control models and technologies. 2002, Monterey, California, USA: ACM Press. 125-134.
- [5]. Wilikens, M., et al. *A context-related authorization and access control method based on RBAC*, in Symposium on Access Control Models and Technologies. 2002. Monterey, California, USA: ACM.
- [6]. D. Sørby, T. B. Røst, Ø. Nytrø, *Empirical Grounding of Guideline Implementation in Cooperative Clinical Care Situations*, AI Techniques in Healthcare: Evidence-based Guidelines and Protocols (Workshop proceedings), Riva del Garda, Italy, 2006.
- [7]. The Asgaard Project, <http://www.asgaard.tuwien.ac.at> (last accessed: November 2007).

Security, Safety, and Related Technology – The Triangle of eHealth Service Provision

Mario SAVASTANO ^{a,1}, Asbjorn HOVSTO ^b, Peter PHAROW ^c and Bernd BLOBEL ^c

^a *IBB, National Research Council of Italy (IBB-CNR)*

^b *ITS Norway, Oslo, Norway*

^c *eHealth Competence Center, Regensburg University Hospital, Regensburg, Germany*

Abstract. The developing of innovative solutions in the emerging eHealth market requires strong economic efforts which may be justified only in presence of particularly suitable boundary conditions. Among the factors retained of primary importance for the development of eHealth, a correct approach to id-management is unanimously considered fundamental. Three keywords in the id-management context appear particularly important: standardization, security and safety. Standardization may contribute to increase the size and duration of the eHealth market, while security and safety may encourage all the stakeholders to trust in a appropriate and safe management of all the very sensitive personal data involved in the eHealth applications. The aim of the present paper is analyzing some security and safety issues in eHealth from the particular prospective of the identity management and standardization. The paper highlights the mission of the EU funded “BioHealth” project whose mission is to increase the stakeholders’ knowledge about existing and emerging standards in eHealth with particular reference to identity management [1].

Keywords: eHealth, Security, Safety, Ethics, Biometrics, RFID, Standardization

Introduction

The healthcare and welfare domains, in both developed and developing countries, are turning towards an extended interaction among different stakeholders. A significant part of this intense communication consists of very sensitive personal information and, therefore, appropriate measures have to be adopted to protect data. If patients are not confident that their data will be acquired, transmitted and stored in a secure and confidential way, they will not be forthright and reveal accurate and complete information. On the other side, if healthcare providers themselves are not confident that the organization responsible for the management of the records will keep them secure and confidential, they will probably limit the disclosure of data. Having assessed that security and privacy represent two fundamental factors in eHealth, there are other issues that play important roles, with particular reference to id-management procedures which represent one of highest level key factors of eHealth.

¹ Corresponding Author: Mario Savastano, IBB-CNR, c/o Department of Electrical Engineering of the Federico II University of Napoli, Via Claudio 21, 80125, Napoli, Italia; E-mail : mario.savastano@unina.it

Many experts suggest that biometrics and RFID (Radio Frequency Identifiers), will find always more a proactive integration with eHealth but, on the other way, some concerns regarding the safety these technologies are raised. For example, in some particular contexts, despite of the large and growing diffusion, iris recognition give sometimes rise to suspects for a potential damage of the eyes and the same RFID identifiers are under investigation because of the electromagnetic pollution. Sticking to the existing safety standards and informing the users of a perfect compliance may represent a valid answer to the concerns.

1. Materials and methods

Advanced concepts of eHealth place the citizens in the focus of a net-centric architecture in which the peripheral is represented by cards and tokens that can bear medical data, improve accessibility to information on services and data, or just serve as authentication token. Since they enable a high-level healthcare data and services access and provision, aspects like security, safety, ethics as well as underlying and supporting technologies need to be clearly addressed before establishing a technical solution. The preferred way to deal with the future challenges of modern healthcare requirements shall be a well-balanced combination of mobile personalized security devices and advanced health networks for both information and services provision. The identity management for all principals (persons, systems, applications, devices, components, etc.) is the key for all further security and safety services. Thus, concepts of security, safety, privacy, quality, and ethics play an important role in the framework of advanced health services. Technologies like biometrics, Radio Frequency Identification (RFID) and Near Field Communication (NFC) are able to technically support the legal, political, and social requirements for such advanced healthcare and welfare service provision [2], [3], [4].

2. Security and safety requirements

The security requirements in the medical area, technically speaking, are not particularly different from those required in other domains. Apart from a very demanding and dynamic privilege management and access control policy, medical and health applications (like in hospital, diagnostic images, laboratory information systems, General Practitioners office software and many other software solutions) base their security functions' provision on available proper mechanism and algorithms for authentication (identification and verification), identity management, confidentiality, integrity as well as availability and accountability [5], [6]. Additionally, aspects related to mechanical, electrical and electromagnetic safety, and play a crucial role in eHealth because safety of patients overrules virtually any other legislation, both in normal and emergency operations. Every technology applied in healthcare needs to strictly and comprehensively address these security and safety requirements derived, e.g., from respective legislation and ethical rules [4].

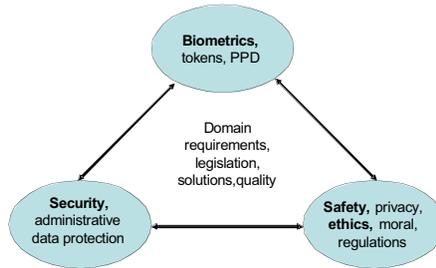


Figure 1. The eHealth “Good Practices” Triangle

2.1. Security requirements towards advanced technologies

Security technologies are frequently used for enabling trustworthy communication and application security services [5]. With reference to Object Management Group (OMG)’s definition of “principal” [7] a basic security principle reflects a certified binding of a principal to its electronic unique identifier or assigned properties, rights and duties, also called attributes of that principal. Application security services deal with authorization and access control to data and functions. Besides, they also cover accountability of principals, audit track and auditing of these principals and services ensuring integrity, confidentiality of data and functions. Communication security services concern the identification, authentication, and verification of communicating principals. In an end-to-end secure communication environment (object security), these services are used for authentication and control of access rights of principals communicating as well as integrity, authenticity, confidentiality, and accountability including non-repudiation of origin, receipt, and information exchanged. For such security objects, security-aware principals are needed [2], [5].

Integrity and confidentiality of communicated data may also be provided at a system level transparent to the application and to the user following, but not requiring, the user’s awareness for those security measures (channel security). Another important requirement for both communication and application security concerns the availability of information and services. More information regarding the different security categories, services, and underlying mechanisms can be found in [5], [6].

2.2. Safety Requirements towards Advanced Technologies

In the development of advanced technologies particular attention should be paid to safety requirements. With reference to the e-health context a specific interest is represented by some authentication devices such as biometric sensors or RFID.

2.3. Safety Requirements in Identity Management

There are two main sources of concerns for safety in the domain of the biometric authentication. The first is represented by the possibility of being infected touching the sensors (e.g. hand-geometry or even fingerprint readers). Even if this possibility may be considered equivalent to that arising in touching a door knob or a telephone, it is anyway true that hospitals represent high-risk locations. Hospital-acquired infections (HAIs), also known as healthcare associated infections, encompass almost all clinically

evident infections that do not originate from a patient's original admitting diagnosis [8]. Nosocomial infections are caused by viral, bacterial, and fungal pathogens. Therefore also the use of a sensor may result in a potential exposure to risk.

Other concerns may arise from the use of biometric sensors which use the eye as a source of information. Iris recognition systems use LED (Light Emitting diodes) which diffuse near-infrared light (NIR) to improve iris detail with dark irises. Unlike UV, IR does not have the energy to produce photochemical damage but NIR illuminators may pose safety issues since the eye does not respond to NIR and does not protect itself as with visible light by means of pupil contraction, avoidance or blinking.

As it attains safety standards or iris recognition systems, LED Eye Safety Standards apply and therefore the following standards and regulations (for references see the respective organizations' homepages):

- ANSI Z136.1 "Safe Use of Lasers"
- American Conference of Government Industrial Hygienists (ACGIH) 'Threshold Limits Values', 1994
- ICNIRP (1996) laser guidelines for exposure limits (ELs)
- IEC / EN 60825-1
- International Commission on Non-Ionizing Radiation Protection. Guidelines on limits for laser radiation of wavelengths between 180 nm and 1,000 nm. Health Phys. Defined in 71:804–819; 1996
- International Commission on Non-Ionizing Radiation Protection. Revision of guidelines on limits for laser radiation of wavelengths between 400 nm and 1,400 nm. Defined in Health Phys. 2000

While biometrics is used for persons, RFIDs have a primary role in e-health since they improve dramatically the identifications and traceability of materials. Some concerns because of the pollution induced and some documents highlight the necessity of further investigation. Main safety standards for RFID are:

- EN55022 Class A equipment (EN61000-6-3:2001) for emissions
- EN61000-6-2:2001 for industrial immunity
- EN60950:2000 safety standard for Information Technology Equipment

3. Technical Challenges and Solutions for Security and Quality

Healthcare does not allow any kind of compromise in terms of confidentiality, integrity, availability, accountability, authenticity or reliability and therefore require an particularly safe management of data since, compromising the rating of a hospital's IT assets, is very likely to have an unfavorable impact, including the risk of a significant financial loss. Consequently, there is an increasing and critical need to protect information and to manage the security of information and communication systems.

While the original motivation for introducing IT security measures has often been security enhancements, appropriate security solutions also offer substantial potential for cost savings and for accomplishing new business opportunities. The ISO/IEC 20000 standard benchmarks the capability of organizations in delivering managed services, measuring service levels and assessing performance. The implementation of ISO/IEC 20000 will reduce operational exposure to risk, meet contractual and tendering requirements, demonstrate service quality and deliver the best possible service. Accordingly it can result in cost savings for users, large or small organizations as well as increased productivity and improved customer service. Regarding software asset management, the implementation of "ISO/IEC 19770-1:2006, Information technology

– Software asset management – Part 1: Processes”, enables organizations to benchmark their capability in delivering services, measuring service levels and assessing performance. Until now the application of these business processes has been arbitrary, and relatively few organizations have been able to implement a comprehensive software asset management strategy with the potential of massive savings in license costs and maintenance fees.

4. Discussion of Results and Strategies

Although everyone recognizes the importance of sticking to standards in the design of eHealth applications, its intrinsic interdisciplinarity represents an evident factor of complexity. Moreover, even if all stakeholders of a project have a sufficient knowledge of the technical and standardization domains, some aspects, such as privacy related or ethical issues are difficult to approach both due to their uncertain boundaries both because of the different perception at the international level. In the course of the BioHealth project, such aspects have been clearly individuated but, at the same time a certain difficulty has emerged in promoting standards in eHealth. Three major issues seem to have priority: (i) make a selection in the domain of stakeholders to find the most appropriate ones for the relevant application, (ii) adopt user-friendly approach, such as multimedia (video, animation, simulation, etc.) to highlight the benefits of standardization, and (iii) propose a centralized approach, at a EU level, to manage the identity management problems concerning technical and ethical issues [3], [4], [5].

5. Conclusion and outlook

A reliable and secure identification is the basis for all advanced security and safety concepts. This is particularly true for health information systems and applications which require an empowerment of all parties (principals) requiring a secure and trustworthy way of communication and collaboration. Moreover they depend strongly on common acceptance which, in its turn, is strictly correlated to privacy and ethical issues [8], [9], [10].

Different technologies including biometrics and RFID, allow for guaranteeing high-level security and safety services addressing proper identification of both human beings and goods but diffusion of standards in these fields is still away from a satisfactory level. Projects such as BioHealth, promoting the knowledge and dissemination of standards, may be extremely useful in supporting eHealth applications but, at the same time, they often require a time frame that exceeds the duration of the project.

Acknowledgement

The authors are in debt to the European Commission for supporting and funding the “BioHealth” project within Europe INNOVA Platform Initiative as well as many other partners and organizations (including ISO TC 215, ISO / IEC JTC1 SC 37, CEN TC 251, CEN TC 224, EFMI, ICAO, CEN/ISSS, IEC, ITU, and HL7) for their permanent support and cooperation in all areas of security, safety, and privacy standardization and policies [1].

References

- [1] EU project BioHealth “Security and Identity Management Standards in eHealth including Biometrics - Specific Requirements having an Impact on the European Society and on Standardization”. Scheduled 2006 - 2008. <http://mirc.gsf.de/biohealth/> (last accessed November 15th, 2007)
- [2] Blobel B, Pharow P (Eds.): Advanced Health Telematics and Telemedicine. The Magdeburg Expert Summit Textbook, pp. 21-28. Series “Studies in Health Technology and Informatics” Vol. 96. IOS Press, Amsterdam 2003
- [3] Kluge EHW: Medical Narratives and Patient Analogs: The Ethical Implications of Electronic Patient Records. *Methods of Information in Medicine* 1999 38 4: 253-259.
- [4] Ball, MJ, Douglas JV: Redefining and Improving Patient Safety. *Methods of Information in Medicine* 2002 41 4: 271-276.
- [5] Blobel B: Analysis, Design and Implementation of Secure and Interoperable Distributed Health Information Systems. Series “Studies in Health Technology and Informatics” Vol. 89. IOS Press, Amsterdam 2002.
- [6] Pharow P, Blobel B: Security Infrastructure Requirements for Electronic Health Cards Communication. In: Engelbrecht R., Geissbuhler A. (Eds.): *Connecting Medical Informatics and Bio-Informatics. Proceedings of MIE 2005*. Series “Studies in Health Technology and Informatics” Vol. 116. IOS Press, Amsterdam 2005.
- [7] <http://www.omg.org> (last accessed November 16th, 2007)
- [8] Report on Hospital Acquired Infections and its Consequences for Patient Safety. <http://www.emedicine.com/ped/topic1619.htm> (last accessed November 13th, 2007)
- [9] Pharow P, Blobel B, Hildebrand C: The Role of Patient Health Cards in an Integrated eHealth Environment. Submitted to STC 2006: Integrating biomedical information - “From e-Cell to e-Patient”. Timisoara, Romania, 2006.
- [10] Engel K, Kuziela H, Blobel B, Pharow P: Security Services for the HemaCAM Project. In: Engelbrecht R, Geissbuhler A (Eds.): *Connecting Medical Informatics and Bio-Informatics. Proceedings of MIE 2005*. Series “Studies in Health Technology and Informatics” Vol. 116. IOS Press, Amsterdam, 2005.

10. Standardization

This page intentionally left blank

Developing a Standard for Personal Health Devices based on 11073

Malcolm CLARKE¹
Brunel University, UK

Abstract. This paper describes the process and outcome of the efforts to develop a new standard for Personal Health Data (PHD) based on the existing 11073 family of standards for medical devices. It identifies the requirements for a standard that is to be applied to small devices with limited resources of processor, memory and power and that will use short range wireless technology. It describes how existing components of 11073, such the Domain Information Model and nomenclature have been used and adapted to create the new standard.

Keywords: Standards, telehealth

Introduction

The 11073 family of standards for medical devices [1, 2, 3] has existed for many years but its use has been limited. This has variously been explained, but that it was designed for plug and play for intensive care unit (ICU) devices has been a major influence. Such devices normally are mains powered, are connected to wired networks and have high quality processing capability. The protocol, based on OSI, is often criticized as being heavyweight and complex. In its current form, it did not appear appropriate to be used as the basis of a new standard for personal health data (PHD) devices. However, with the expected rapid increase in the demand for health devices in the home with capability to communicate results, a standard capable of operating in this environment is essential.

The 11073 family of standards is partitioned into a set of standards covering the many aspects of communicating the semantics of medical data from device to manager. This includes a Domain Information Model (DIM), nomenclature, device specializations, device behavior, communication transports, and communication protocol. The 11073 family of standards has also acted as an umbrella for medical device standards.

¹ Corresponding Author: Malcolm Clarke, Brunel University, Uxbridge, Middlesex, UB8 3PH, UK; E-mail: malcolm.clarke@brunel.ac.uk

1. Background

1.1 The IEEE 11073 PHD Working Group

The IEEE 11073 PHD Working Group (WG) was established to develop a new medical standard that would be used for the typical PHD device. It was accepted that any new standard would need to be implemented within the limited resources of such devices. It was also expected that the standard would align with current developments in the Bluetooth SIG and USB SIG to develop health profiles. The work group has set the task to develop a common base protocol that will work with an initial set of seven device specializations (pulse oximeter, pulse/heart rate, blood pressure, thermometer, weighing scale, health/fitness and blood glucose).

1.2. The Standards Process

Initially four proposals were submitted to the group for consideration as a basis for the standard. However no one proposal satisfied all the requirements and a process to develop a combined proposal was adopted. This included identifying a template of a set of minimum requirements, a set of preferred requirements, and a set of mandatory behavior. Proposals were compared and strengths of each identified to inform the final proposal.

The final proposal was mainly based on the 11073 standard but included important changes to accommodate the resource requirements of PHD devices and to incorporate advantageous characteristics of the other protocols.

2. The Protocol

2.1. Protocol Overview

The work of the IEEE PHD 11073 is defined by the framework as shown in figure 1. The overall task is concerned with defining the protocol at layer 7. The transports which provide layers 1-6 are defined elsewhere and are outside the scope of the work of the group, although there is close liaison with groups such as Bluetooth SIG to ensure compatibility. However note that the group has set an objective to make the protocol transport agnostic.

The existing 11073 standard uses OSI layer 7 and utilizes existing functionality of CMISE and ROSE. It was quickly apparent that an optimized exchange protocol was required. This would provide the same functionality as OSI layer 7, but implement it in a lightweight fashion, eliminating any redundant features, and be fully defined in the new standard 20601, which would simplify implementation.

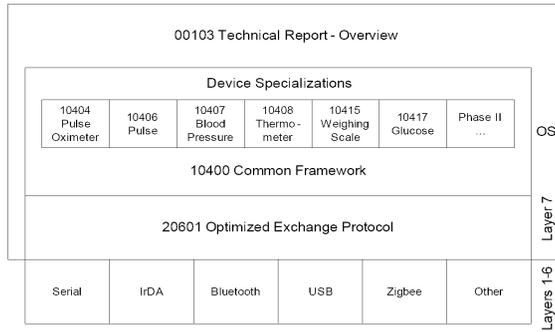


Figure 1. Overview of the IEEE PHD 11073 Framework.

2.2. Domain Information Model (DIM)

The existing DIM of 11073 was used, but it was simplified for PHD devices by constraining the scope of the model and restricting and flattening the hierarchy. Abstract Syntax Notation (ASN.1) is used to describe the model, and this may also form the basis of definitions of data structures for other languages.

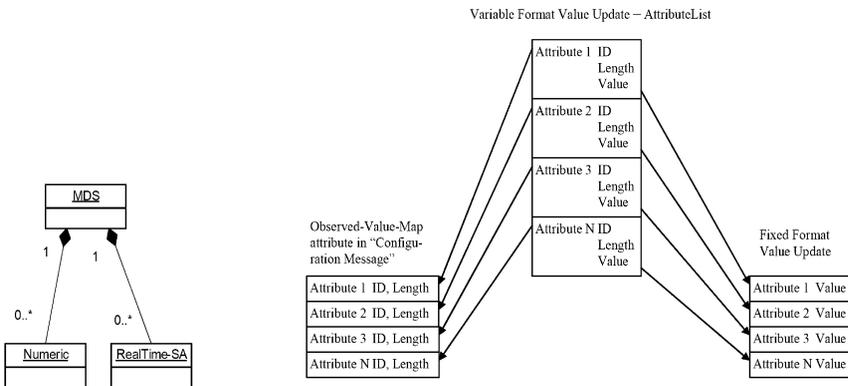


Figure 2. Optimized PHD DIM. Figure 3. Fixed and Variable Data Formats.

The current optimized DIM for the PHD has five objects: the Medical Device System (MDS); the Numeric; the Real Time Sampled Array (RT-SA); the scanner; and the persistent metric store (PM-Store). The MDS has as attributes all the information pertaining to the device and its operational status, such as unique device ID, device

configuration, and time functions. Attributes also contain product specification in text form. Attributes may be determined by using the GET method defined in 20601.

Numeric objects relate to the physiological parameters and have as attributes the mechanism to obtain an observed value and its status such as units and the timestamp. The numeric object is defined to permit intermittent observations to be reported. Observations may be reported using four methods: the manager may make a specific request for currently available data; the manager may request data to be reported as they become available for a specified time; the manager may request data to be reported as they become available for an unbounded period of time; the agent may send an unsolicited observation. The RT-SA is optimized to report an array of observed values as a single data transmission, which reduces protocol overhead and would be used for real time streams with high data rate and requiring low latency, such as the plesythmogram.

The protocol is further optimized by allowing for fixed and variable format of data transmission (figure 3). In variable format, each observation carries its attribute ID, the length of the entry and the numeric value. If a stream of observations is established, each having the same attributes, then the common attributes can be defined in advance of the transmission so that only the values need to be reported each time. This is common attribute list is defined as the Observed-Value-Map is applied to each set of values reported and will reduce the transmission burden. The idea is further extended to the concept of defining standard devices with standard configuration. In this case there may be no need to define the Observed-Value-Map in advance, so reducing transmission burden during association further. A device may define itself as supporting extended functionality and use the variable format to allow flexibility.

The scanner object provides an optimized mechanism to report the observed values and attributes of several objects. It extends further the concepts of the attribute map of figure 3 to include the attribute maps of several objects in an optimized format.

The PM-Store has been provided to store large amounts of data whenever the agent operates without connection for later retrieval. To suit simple agents, the PM-Store is designed with only two levels of hierarchy, so that there can be multiple PM-Segments within multiple PM-Stores. This simple model should be able to model most practical situations. The PM-Segment is accessed to retrieve the actual data using PM-Store actions. Each PM-Segment contains a homogeneous set of elements, and multiple PM-Segments may be used to model and store different aspects of the data of a device.

The Medical Device Encoding Rules (MDER) [3] are used to convert ASN.1 structures to binary transmissions. Although essentially the same as DER, they apply some optimizations to the protocol by having fixed size coding and removing some of the features, and so aligns with the needs of PHD devices. Note that network byte is big endian.

2.3. Communication Model

The communication layers have been assumed to be a point to point link and provide the mechanism to transfer data from peer to peer device. It is further assumed that whenever the transport indicates a connection, the state machine (figure 4) moves to the connected state and the agent is placed in the unassociated state. The agent will initiate the association between itself and the manager by issuing an association request and will enter the associating state. The association request will contain configuration ID and allow the manager to determine if it should accept the request and if it already has

configuration information from an earlier association. If the manager does not have configuration information it must request that configuration information is sent by the agent prior to entering the operating state. The configuration information sent by the agent will include information on the objects in the device and a handle number by which they may be accessed. This step is bypassed if the configuration is already known and assumed unchanged. Manager and agent will then enter the operating state. An association release and its response will take manager and agent back to the unassociated state, and this is the preferred method to disconnect devices.

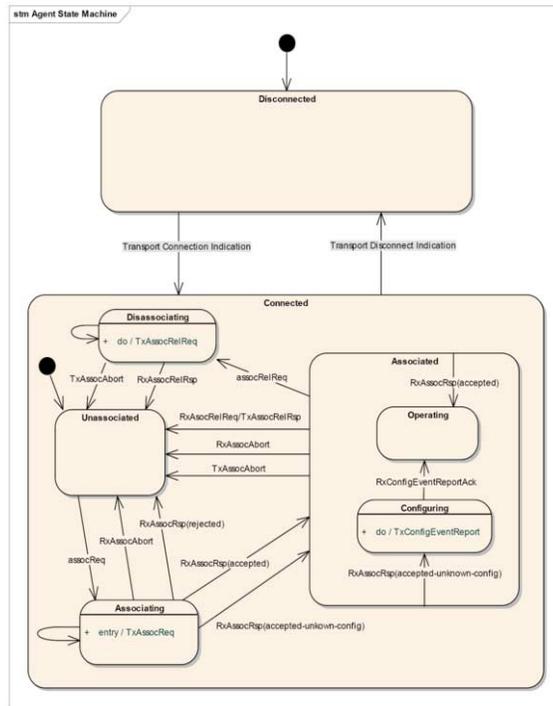


Figure 4. Association State Machine.

3. Conclusion

The IEEE PHD WG has focused on producing a protocol for medical devices that is optimized for low capability agents that have limited resources of processing power, memory and power for communication. It has reduced the size and complexity of the existing 11073 standard by reducing the data transmission sizes through defining a lightweight application layer, removing the session and presentation layers of OSI, and making assumptions of the transport layer. The DIM has been constrained and its hierarchy flattened to create simplified models more appropriate to PHD devices. An

optimized reconnection protocol can remove the need to transmit the configuration of an agent already known to a manager or for standard devices. The protocol aligns with the existing DIM and utilizes the existing nomenclature to leverage the 11073 standards and to provide a framework for extensibility. At this time the standards are currently draft status.

References

- ISO/IEEE 11073-10101: Health informatics — Point-of-care medical device communication — Part 10101: Nomenclature. First edition 2004-12-15.
- ISO/IEEE 11073-10201: Health informatics — Point-of-care medical device communication — Part 10201: Domain information model. First edition 2004-12-15.
- ISO/IEEE 11073-20101: Health informatics — Point-of-care medical device communication — Part 20101: Application Profiles – Base Standard. First edition 2004-12-15.

SOA approach for integration of departmental systems

Timo ITÄLÄ^{a1}, Jari UKKOLA^b, Aino VIRTANEN^b, Juha MYKKÄNEN^c

^a *Conceptia Oy, Helsinki, Finland*

^b *Hospital District of Satakunta, Pori, Finland*

^c *University of Kuopio, HIS R&D Unit, Kuopio, Finland*

Abstract. In this paper a unified method for integration of departmental systems into the main systems of a healthcare organization is described. The approach is based on combining Business Process Modeling (BPM) and Service Oriented Architecture (SOA) methods and technologies. A top-down approach is used for modeling the care process and supporting care services which in turn are decomposed down to such a level of granularity that they can be described and implemented as web services described with Web Services Description Language (WSDL) documents. Then a bottom-up approach is used for wrapping the existing departmental systems and their interfaces into web services using Enterprise Service Bus (ESB). Finally the orchestration of the services is described using executable Business Process Execution Language (BPEL) code.

Keywords. Modeling, Process, EPR-CPR-EMR, Interfaces, Systems architecture, Clinical information systems, Application Integration, Service Oriented Architecture, Enterprise Service Bus

Introduction

The core applications of most healthcare organizations are patient administration systems (PAS) and electronic patient record (EPR) systems. In addition to them, applications which are used within clinical departments for some specific purposes such as diagnostics are increasingly and rapidly emerging. These applications need patients' demographic data and they produce resulting data of examinations performed to the patients. Today most of these departmental systems are not well integrated into the PAS or EPR resulting in a need to re-enter the data or inability to share the resulting notes and images in digital format. Healthcare organizations need to integrate the departmental systems with the core systems of the organizations in order to support the care processes and flow of information. The integration solution should be reusable and unified to save costs and to speed up the integration efforts. Approaches such as Health

¹Corresponding Author: Timo Itälä, Conceptia Oy, Puotilantie 17, 00910 Helsinki, Finland, E-mail: timo.itala@conceptia.fi

Level Seven (HL7) messages and Clinical Document Architecture (CDA) documents [1] are often not used in a uniform way or their focus has been to integrate applications between organizations.

1. Methods

This work was conducted to support the emergent integration needs of departmental systems within the hospital district of Satakunta in Finland in collaboration with the SerAPI project of University of Kuopio [2, 3]. The practical approach taken for this task combined a top-down modeling of care process and service processes based on Service Oriented Architecture (SOA) [4] with a bottom-up approach which wraps the applications to web services using an Enterprise Service Bus (ESB) [5]. Such an approach is necessary to enable the improvement of care processes while also utilizing the information and functionality of existing systems [3].

2. Results

2.1. Workflow of a department using a departmental system

The objective was to develop a unified approach for the integration of departmental systems in the hospital. The endoscopy department and their Endobase system were chosen as the pilot project. More than ten other systems to be integrated were also analyzed, such as Retinal Image Screening, Dermatology, Audiogram, ECG etc. The common need for all the departmental systems was identified: The patient data should come automatically from the PAS to the departmental system and the resulting data of the examination should go automatically from the departmental system into the EPR.

The conclusion was that there is a need for an abstraction layer on top of the actual systems which would hide their technical details and give a unified interface for transferring patient data from PAS into the departmental system and resulting data out of the departmental system into EPR.

2.2. Care processes and care services

A service in SOA can be a process, a sub-process or a process step [4]. The aim was to model and decompose the processes and services down until such a level of granularity that the services can be implemented as web services [3]. Using this top-down approach the process depicted in Figure 1 was modeled.

In the model the care process is the most central process of the hospital, and contains only such steps that are related to the actual care. The model is a generic one and covers all the different specialties and patient cases [6]. The customer of the process is the patient. The owner of the process is the doctor in charge of the unit. The instance of the process (an episode of care) is triggered by an event which is the arrival of a referral or the arrival of the patient for emergency care.

The care process is a consumer of several care services which can be invoked in any step within the care process. The service requests are made by the doctor in charge of the patient. Typical services are diagnostic tests or examinations such as lab tests,

radiology examinations, endoscopy examinations etc. Typical services are also orders for treatment such as medication, surgery operations, therapies etc. Not so evident services include logistical arrangements such as visit to an outpatient clinic, episode within a ward or an appointment for an endoscopy examination. In this last case the requestor of the service may be actually the same doctor as the provider of the service.

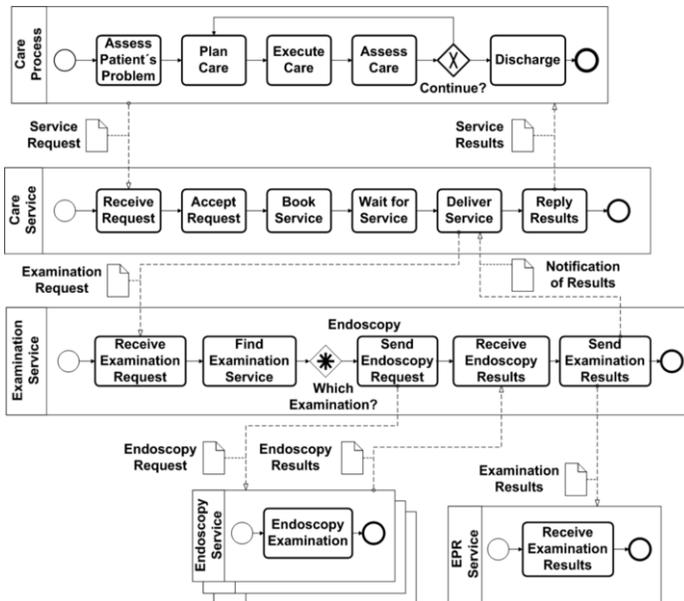


Figure 1. Care process and care services

These care services consist of process steps which are executed according to a predefined logic and which may invoke other services. An example of a service which is relevant for the purpose of integrating departmental systems is an *examination service*. In the abstract model all requests for examinations are served by this examination service. In the example it invokes the endoscopy service although all the other examinations may be invoked in a similar fashion. The endoscopy service returns the results of the examination which in turn are delivered to the electronic patient record by invoking the EPR service.

2.3. Modeling Web Services

The model has reached the level of granularity where identifying potential web services, their operations and messages can be started. Potential web services are shown in Figure 2 using an interaction diagram.

A web service (box) in Figure 2 corresponds to a pool in the BPMN model. A web service contains operations which are shown beneath the service. An operation sends a message to another service or receives a message from another service as described with the arrows between the services.

The web services were described using WSDL (Web Services Description Language) documents. WSDL defines abstractions of services, their operations and messages and a binding to a concrete protocol and network address as an endpoint providing the services.

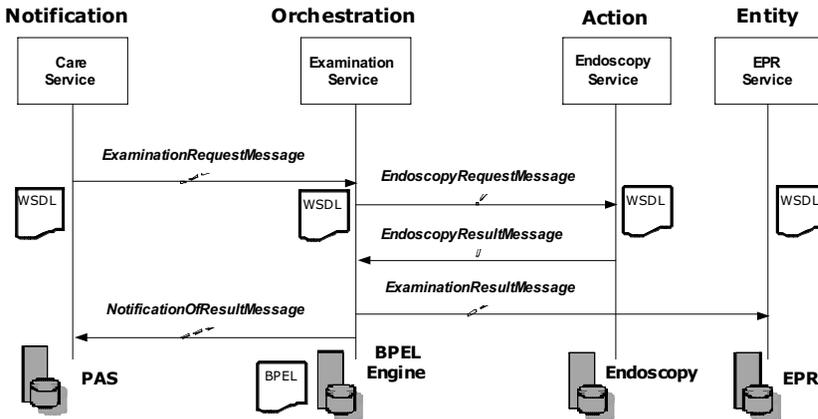


Figure 2. Web Services

On this level different types of services and their realizations using different applications and systems can be identified.

- Care Service is a *notification* type service. It is implemented on top of the existing PAS so that whenever a request for an examination is entered into the PAS this service will send a notification message containing that request.
- EPR Service is an *entity-centric* service. It provides operations for reading, querying and managing patient records. The example displays only one one-way operation for receiving result data from examinations to be included in the EPR. These services are implemented on top of the existing EPR system.
- Endoscopy Service is a *task-centric* service which is implemented on top of the existing Endoscopy application. It receives requests for examinations and asynchronously returns their results when they are ready. Modeling the endoscopy system as a web service helps to avoid the double data entry which is done today. Other departmental systems can be modeled in a similar fashion to provide a unified way for their integration.
- Examination Service is a service whose function is to *orchestrate* the invocation of other services in order to carry out the examination and store its results. This orchestration service is implemented with BPEL (Business Process Execution Language) code executed at run time using a BPEL Engine.

BPEL code defines the partners and their roles participating in the orchestration and also the process execution logic. A modeling tool was used to generate the initial BPEL code based on the WSDL descriptions and the straightforward logic of the Examination Service process.

2.4. Implementing Web Services

After modeling a unified way to send the request for an examination and related patient data to the departmental system and later to receive the results of that examination, there is one remaining question to be solved: How can the differences in protocols and formats of the various departmental systems be hidden? A middleware solution which will help us in this question is called the ESB (Enterprise Service Bus) [5]. In this case it provides the following functionality:

1. It provides the web service endpoints described with the WSDL documents using the addressed locations, operations and messages.
2. It has an intelligent routing capability between the web service address and the physical address of the actual system.
3. It connects to the respective systems using their native protocols and message formats, such as HL7 V2.
4. It provides message content transformations between XML and different other formats, such as HL7 V2.

A logical structure of the ESB is shown in Figure 3. In the setup it consists of an Intersystems Ensemble message broker and transformation services needed to abstract the native interfaces of different departmental systems into unified services described with the WSDL documents.

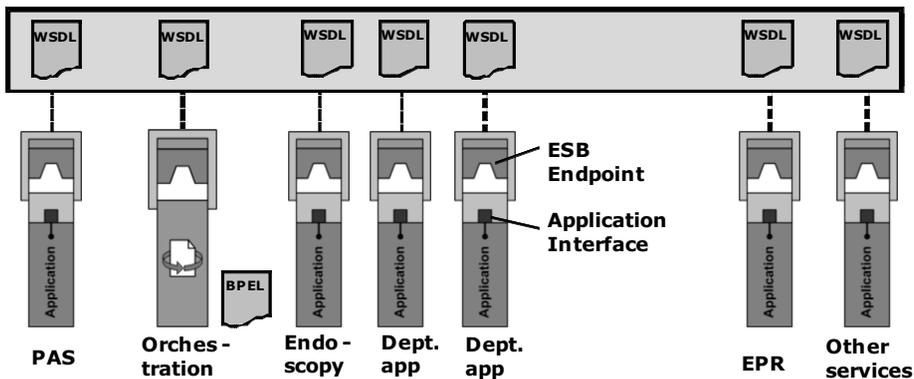


Figure 3. Enterprise Service Bus (adapted from [5])

3. Discussion

The departmental systems were abstracted to be providers of unified web services so that the integration of those systems can be done in a unified way using their native interfaces. The essential idea was to make a clear distinction between the abstract care process and supporting care services, which in turn may be consumers of more fine grained services.

Also a conceptual model and data model were defined to provide unified reference model for holistic information integration. The ideal solution would be that the ESB communicates using unified XML data formats, unique identifiers for all entities and common code sets and vocabularies. The examination results could be translated into nationally used HL7 CDA R2 formats [2] using capabilities provided by the ESB.

The modeling of care processes on higher level using care services may lead to discover quite new ways on organizing care using networks of different service providers. In general, this kind of abstraction and architecture forms a basis for systems which can adapt to inevitable structural changes in health services delivery.

This generalization, however, also poses challenges to the modeling and implementation: as the generic requests are not delivered directly to the specific service, the service-specific models need to be supported by the process controller or the generic request formats. In addition, directory services could make logical

deductions and choose a proper examination service based on data entered into the PAS. The definition and division of these responsibilities in a highly distributed architecture is not a straightforward task.

Another challenge is related to the acquisition of systems: There is a need for vendors who would be responsible for the orchestration and integration of the services. In addition, the end-to-end testing of the care processes and services so that they produce the expected results in all situations need further research.

In relation to standardization, this work aligns with SOA-driven standardization efforts for healthcare (such as Healthcare Services Specification Project [3] where the authors have been involved). There are also other service-labelled healthcare IT standards such as CEN HISA which will be assessed in relation to the SOA principles [4] in the future.

The generic process model of the approach also supports generic EHR models such as [6]. Similar service-oriented approaches have been used for patient portals [7] and integration between clinical and public health information systems [8], but without considering the enterprise service bus as an enabler.

4. Conclusions

Service oriented architecture gives a new approach for connecting different systems in a standardized way, although they may have their own proprietary interfaces. An abstraction of the care process and supporting care services hide the richness of different patient cases and combinations of how care is actually delivered. This abstraction and categorization of services are essential to promote reuse and utilization of generic models and infrastructures and to prevent complex and costly point-to-point definition of solutions. The benefits of a unified and reusable integration solution are reduced amount of work, cost savings and more rapid implementation. In the future Satakunta Health District is considering using this approach for integration of all new departmental systems.

References

- [1] Mikola T, Wanne O, Itälä T, Virtanen A. Implementing Regional Information System for Social and Health Care in Finland, Satakunta. Surján G, Engelbrecht R, McNair P (eds): Health Data in the Information Society. Proceedings of MIE2002, IOS Press, 2002, pp. 723-728.
- [2] Mykkänen J, Korpela M, Ripatti S, Rannanheimo J, Sorri J. Local, Regional and National Interoperability in Hospital-Level Systems Architecture. *Meth Inf Med* 2007;46:470-475.
- [3] Mykkänen J, Riekkinen A, Sormunen M, Karhunen H, Laitinen P. Designing web services in Health Information Systems: from process to application level. *Int J Med Inf* 2007;76(2-3):89-95.
- [4] Erl T. Service-Oriented Architecture, Concepts, Technology and Design, Prentice-Hall, 2005
- [5] Chappel DA. Enterprise Service Bus, O'Reilly, 2004
- [6] Bernstein K, Bruun-Rasmussen M, Vingtoft S, Andersen SK, Nohr C. Modelling and implementing electronic health records in Denmark. *Int J Med Inf* 2005;74(2-4):213-220.
- [7] Duncavage S, Mathe J, Werner J, Malin BA, Ledeczi A, Sztipanovits J. A Modeling Environment for Patient Portals. Proceedings of the 2007 American Medical Informatics Association Annual Symposium, AMIA, 2007, forthcoming.
- [8] Lopez DM, Blobel BGME. Connecting Public Health and Clinical Information Systems by Using a Standardised Methodology. Kuhn K, Warren J, Leong T-Y, eds. Medinfo 2007, IOS Press, 2007, pp. 132-136.

Procurement of prescriber support systems

Karin KAJBJER¹,

Sentensia Q, AB, Stockholm, Sweden

Abstract: Supporting the process of medication selection and electronic management of prescriptions is a high priority issue in the eHealth strategies of many countries today. Procuring such systems can be quite difficult, especially if one should encourage suppliers from different countries to participate. The new ISO Technical Report 22790 [1] provides a new approach to facilitate this process by giving an international basis for specifying the functional characteristics desired. The paper describes the content of the report and discusses the procurement process in the light of the European public procurement directive and patient safety.

Keywords: procurement, prescriber support systems, decision support, ISO, management, standards, patient safety, risk management, EU-directives

Introduction

Health care is a complex system with high quality and safety demands, hence the need to be supported by high quality information systems.

Medication is an effective means of improving health though the use of medication is costly and introduce risks to patient safety. Many countries have listed information systems to improve the processes related to prescribing as a top priority for health IT. In order to get the best products at the lowest possible cost, it is important that the procurement of such systems can work effectively on an international or at least common European market. In the case of a publicly financed entity, procurement of IT support as other things must in the EU comply with the common legislation aiming at providing equal opportunities for industrial suppliers mainly SMEs from all countries. Also, for completely private health care organizations and restricting oneself to one country, procurement of complex IT systems is not easy.

A procurer may need to spend a lot of time and costs to define and express all the requirements on the product which for international competition may need translation to usually English (at least in the small countries). The suppliers need to spend a lot of time to try to understand and respond to the differently formulated requirements from each customer. In the end many functional features are described by the suppliers in their own jargon and comparisons of the different offers is a difficult task.

In many other fields, there has been a long term effort to develop technical standards that defines various product characteristics in order to facilitate the

¹ Address for correspondence: karin.kajbjer@telia.com

procurement process, for both suppliers and customers. However, in the rapidly changing eHealth field such standards are rare. The great differences between countries regarding information on medicinal products and various degrees of maturity and strategies for national eHealth systems has made it difficult to develop international standards defining agreed requirements for all the relevant aspects of prescriber support systems.

ISO, the International Organization for Standardization has developed the new Technical Report ISO/TR 22790 Health informatics - Functional characteristics of prescriber support systems. This is not an international standard with normative provisions. It takes another approach. This *informative* document provides an agreed description of the various functionalities and information used in a common terminology which is intended to be helpful for procurement processes. The procuring organization (and possibly national bodies) will select among the many possible requirements from the ISO/TR according to local prerequisites and priorities. The advantage is that it is much easier to choose from existing well formulated requirements and most importantly, the suppliers also international ones can be familiar with all expressed requirements which speeds up and adds quality to the process. It has also been an objective of the ISO group that some requirements that are formulated here will provide important inspiration for suppliers to develop better products and for procuring organizations to demand functionality they may not have considered.

1. Methods

This study is based on a literature study of several basic documents on public procurement legislation in the European Union:

(http://ec.europa.eu/internal_market/publicprocurement/legislation_en.htm).and in particular the "Public Procurement Directive [2].

It is also based on the legislative situation as regards Medical Devices in the EU (<http://ec.europa.eu/enterprise/newapproach/standardization/harmstds/reflist/meddevic.html>) including the basic Directive 93/42 [3].

Regarding the standards development, the author has participated in the relevant working groups of SIS and of ISO/TC 215.

The development of the standard for Functional characteristics started in Sweden in 2001 with a working group for Medication under the Health informatics committee and SIS, the Swedish Standards Institute. The results were published as SIS/TR 2 in the Swedish language in 2003.

In the ISO/TC 215 Health informatics committee, a working group 6 for medication related issues was created and it was proposed to start the work towards an international publication on the same subject. This was approved in 2005 and the Swedish document was translated and much enhanced in numerous discussions in the working group with experts from Europe, Japan, Korea, Canada, Australia and the USA and following the international formal ballot and review during 2006. The revised final version was agreed in the meeting in Montreal in March 2007 and after review and final editorial corrections published by the ISO central office in Geneva, November 2007 in both French and English languages.

2. Results

2.1. The European legislation and prescriber support systems

The general provisions of the public procurement directive [2] has been implemented in national legislations within the European Union and apply when publicly funded health care institutions procure IT systems for prescriptions. These provisions aim to ensure that there is a fair an open market ensuring that suppliers from all countries should be able to offer their products. This is partly ensured by stipulating rules for the technical specifications in a call for tender that preferably refer to European standards or in their absence to international standards. The technical report from ISO is not mandatory to use yet but it is in line with the general principle to use such documents to facilitate the understanding of the requirements.

In many areas which affect safety of persons, there are also other common European legislative requirements applicable for procurement and then not restricted to public bodies. In this field of information systems for health care use, the Medical Devices directive [3] applies and depending on the class of device various provisions apply and in many cases detailed technical standards from the European Committee for Standardization CEN or CENELEC in the electrotechnical field applies. There has been an ongoing discussion between the member states and the European Commission on how to classify and control the market for information systems that have no direct or indirect contact with the patient but still are used for health care purposes. In at least the first ten years of the life of this directive, such systems were not regulated. However, in recent years there has been a change of opinion gradually and even if it is at present unclear exactly how, the interpretation in many member states is moving to include products like prescriber support systems in the controlled area. In view of this it is quite possible that the application of reference documents like the ISO/TR 22790 will become mandatory in a not so distant future. In this year of 2008, the European standard organizations are developing a new work plan in response to a formal mandate from the European Commission to develop new standards as required for eHealth. Prescriber support systems are likely to be considered in this context. However, at this point in time the ISO/TR should be considered a valuable possible help for procurers but without formal requirements to use it from a European legal point of view.

2.2. The ISO/TR 22790 approach

The technical report ISO/TR 22790 from 2007 is intended to be used in the interaction between procurers and suppliers of systems to support the prescriber in all the different aspects. It provides a common conceptual model of information management related to the process of prescribing or ordering medication. The report provides a set of optional business requirements that could be selected by the buyer in a procurement process to be responded to by a tendering supplier. The report does not provide any mandatory requirements but as an informative document gives a common expression of various possible functions meeting different objectives for the health care system.

This document is intended to be used as a guide for a specific organization in formulating and prioritising a subset of characteristics tailored to national or local needs. The complete list of requirements of the ISO report is thus not intended to be a minimum set of requirements that all systems must comply with.

This report contains an introduction to the necessary concepts with agreed definitions and recommended terms and an overview of the relationships between different actors and information flows. The report also provides a functional model based on the objectives of the health care system:

2.2.1. Assessing the patient's need for medication

The main method is using access to Electronic health records

2.2.2. Selecting a medication that can give an optimal result for the patient and current problem

A system can facilitate the use of Clinical guidelines for diagnosis and therapy recommendations. Systems may also provide advice on dosage and information on risks for adverse effects

2.2.3. Making cost conscious selections that can contribute to the cost containment of the insurance or publicly funded health care system as well as patient costs

This includes prices and comparisons of prices and recommendations of local drug committees. Also the consequences of the regulations for reimbursement Examples: Positive list of allowed products, Negative list of disallowed, Reference pricing information or fixed pricing information. Also selection of packages sizes to minimize costs

2.2.4. Issuing a complete prescription in a time efficient manner

E.g. Easy renewal from medication history and using templates for complete prescriptions or dosage only

2.2.5. Transfer the information to a pharmacy

Via printed paper form

Patient data card

EDI via structured messages or other form of direct electronic communication. This can be direct to a pharmacy or via central prescription store or relaying agent

2.2.6. Communicating with the patient

Printing of patient medication list. Electronic communication

2.2.7. Communicating the medication orders to other health care professionals

Process support to medication administration and shared medication histories

2.2.8. Follow up

Functions to periodically follow up the total prescribing by the prescriber and/or unit in this system

2.2.9. Information resources needed to achieve the requirements

The TR defines a set of resources (databases) needed for the prescriber support systems to work. Many of these could be considered as part of a national information infrastructure that should be put in place to allow development and operation of such systems.

2.2.10. List of detailed characteristics to select from in a procurement process

This is the core of the ISO/TR 22790 and for each of the 90 plus requirements it is intended that the procurer indicate (in a call for tender document) how the customer is viewing each of these requirements:

- Not important = 0
- Desirable = 1
- Shall be available later at a defined time = 2
- Shall be available at delivery = 3

In the intended use of the standard the supplier indicate for each requirement the availability of the desired characteristics as:

- The function is not available = 0
- The function can be delivered at a defined time to be specified = 2
- The function is available at delivery = 3

3. Discussion

Standards from ISO and other formal standards bodies such as CEN, the European Committee for Standardization are used in many sectors to facilitate procurement including cross-border trade. In eHealth the major focus on standardization activities has been on interoperability specifications and they certainly have the potential to be of great importance although much of this potential has not yet been transferred to real implementations. This partly because many purchasing organizations have been unaware of them and also due to the fact that some interoperability standards require additional specification to be made in the form of a profile, often for national use, restricting optionalities and making the specification more concrete and hence allowing testing of compliance of products. A study performed for the European Commission during the eHealth Standardization Focus Group [4] demonstrated that many stakeholders are unaware of the fact that the current legislation in Europe actually mandates the reference to primarily European but in the absence of such, international standards in public procurement.

The new ISO/TR 22790 is one of a few international standards that address functional characteristics that are not only related to interoperability but also serving to provide optimal quality of the system used locally. The ISO/TS 18308 [5] is another example but the procurement aspect was not in the main focus of this publication as with the present standard report on prescriber support systems.

There are many advantages to provide some degree of common international framework as in this new publication even in a situation where there are strong differences between countries. It can lead to a more sound market situation and also be

the basis for further work also on a European or International level aiming to go further in defining minimum characteristics.

The legislation on medical devices in Europe, the directive from 1996 has been considered to exclude requirements on much of the eHealth systems that are not directly connected to a device which directly influences the human body. However, this directive is currently under revision and governments are planning for a much increased control of all types of medical software in order to improve patient safety. It will be necessary to support such efforts with specific requirements in technical standards for different types of systems. We will most probably see an increased number of such standards in a near future but the whole process of improving software quality will take time. It also requires the building of appropriate certification schemes ones the basic standards to test against will be present. In a first round most actors agree that a system with self declaration by the market authorization holder against an established formal standard is much better than no clear statements. This does not preclude that in some areas third party testing and certification may be needed.

It is sometimes claimed that publications from standards organizations are complicated and too technical to be possible to understand, especially by health care professionals participating in a procurement process. This is clearly not the case for the specific technical report described in this paper. This technical report is a means of facilitation and quality assurance of the procurement process. The report has been developed by a working group within ISO representing different relevant stakeholders. The procurement of IT-systems for health is often regarded as a business to be handled by the engineers. Using tools like this report, this does not need to be the case but active health professionals can easily understand and express their requirements. Other technical interoperability specifications will need to support some of the claims.

So far there is very little experience to report on actually using this new standard since it was published only a few days ago. Hopefully we will gain more experience from different countries in a near future.

4. Conclusion

A document such as the ISO/TR 22790 on Functional characteristics of prescriber support systems can be a great facilitator of procurement of such systems, contribute to the better functional of the international market and provide a basis for developing better and better solutions for enhanced patient safety in relation to medication.

References

- [1] ISO/TR 22790 Health informatics – Functional characteristics of prescriber support systems. International Organization for Standardization, Geneva 2007.
- [2] Directive 2004/18/EC of the European Parliament and of the Council of 31 March 2004 on the coordination of procedures for the award of public works contracts, public supply contracts and public service contracts (30.04.2004)
- [3] Council Directive 93/42/EEC of 14 June 1993 concerning medical devices
- [4] Report from the CEN/ISSS e-Health Standardization Focus Group, Current and future standardization issues in the e-Health domain: Achieving interoperability.
- [5] ISO/TR 18308. Health informatics -- Requirements for an electronic health record architecture

Enhanced Semantic Interpretability by HealthCare Standards Profiling

Diego M. LOPEZ¹ and Bernd G.M.E. BLOBEL

eHealth Competence Center, Regensburg University Hospital, Regensburg, Germany

Abstract. Several current healthcare standards support semantic interoperability. These standards are far to be completely adopted in health information system development, however. The objective of this paper is to provide a method and necessary tooling for reusing healthcare standards by exploiting the extensibility mechanisms of UML, by that way supporting the development of semantically interoperable systems and components. The method identifies first the models and tasks in the software development process in which health care standards can be reused. Then, the selected standard is formalized as a UML profile. Finally that profile is applied to system models, annotating them with the standard semantics. The supporting tools are Eclipse-based UML modeling tools. The method is integrated into a comprehensive framework for health information systems development. The feasibility of the approach is exemplified by a scenario reusing HL7 RIM and DIMs specifications. The approach presented is also applicable for harmonizing different standard specifications.

Keywords: Standards, semantic interoperability, profiles, reusability, UML, HL7

Introduction

Especially in the context of long-term usable eHealth applications such as Electronic Health Record (EHR) systems, at least two basic requisites for semantic interoperability have to be met: i) agreement on a standardized set of domain-specific conceptual models, and ii) agreement on standardized terminologies associated with controlled vocabularies [1].

Several healthcare standards defining domain models currently exist, such as ISO HL7 21731 “Health informatics- HL7 version 3 Reference Information Model” (RIM), ISO EN 13606 “Health Informatics –EHR communication”, OpenEHR Archetypes, etc. These standards are far to be completely harmonized and adopted, however. Apart from political reasons, lack of tooling, methodologies, bridging between the domains involved, guidelines, completeness, conformance mechanisms, harmonization, etc., are principal barriers for their adoption.

Commonly, Standards Development Organizations (SDOs) use their own notation, and underlying development paradigm to describe their standards. UML, the de-facto industry standard for software modeling, has been used for many of them as language to formally represent the standards semantics. Using UML provides several advantages

¹ Corresponding Author: Diego M. Lopez, MSc, eHealth Competence Center, Regensburg University Hospital, Franz-Josef-Strauss-Allee 11, D-93053, Regensburg, Germany; Email: diego.lopez@ehealth-cc.de

such as tooling support, widely spread use, graphical notation, exchangeability, extensibility, code generation, etc. Unfortunately, UML has not been consistently used in healthcare standards development and – without appropriate adaptation – presents some limitations to describe specific domain concepts.

The objective of this paper is to provide a method for reusing healthcare standards by exploiting the extensibility mechanisms of UML, by that way supporting the development of semantically interoperable systems and components. If systems were based on standard information models (e.g. specifying the platform independent model), this model would be easily transformed into different implementation platforms supported by the Model Driven Architecture (MDA) approach. The resulting systems and components implement the information structures and behavior described in the standard. As a result semantic interoperability at functional and services level is supported.

1. Definitions and Methodology

Profiling in software engineering is commonly defined as a mechanism for extending existing meta-models to adapt them for different purposes. In health informatics, the term profile has been extensively used in different ways. HL7 for example uses profiles to constrain and extend HL7 specifications such as the RIM, Domain Information Models (DIMs), services specifications, etc. The objective of profiling HL7 specifications is to use them in particular environments and/or defined community of users (e.g. specific requirements in some countries). The IHE (Integrating the Healthcare Enterprise) initiative defines Integration Profiles to describe the solution to a specific integration problem (use case) in different domains, e.g. IHE profiles for cardiology, radiology, IT infrastructure, etc., thereby reusing existing standards. OpenEHR defines the Archetype Profile which is a mechanism to specify custom Archetypes for certain kind of information models.

The concept of profiling deployed in this paper has been defined in the UML standard [2] as a formal mechanism to specialize a reference meta-model in such a way that it is possible to adapt that model to a specific platform or domain. Profiles are defined as UML packages which contain stereotypes representing the meta-classes of the meta-model that can be extended. The stereotype is represented as a UML class along with its attributes, and is represented adding the tag <<stereotype>>> before the class name or using a customized shape (see class diagram in Figure 2). A profile is formally defined as a UML model enabling to interchange profiles between tools (using XMI), together with the models having been applied.

The method proposed to support the development of Health Information Systems (HIS) based on healthcare standards, is supported through the Health Information Systems Development Framework (HIS-DF) [3]. HIS-DF is a comprehensive and customizable methodology for the architectural analysis of HIS based on the Generic Component Model (GCM) and deployed as a specialization of the Rational Unified Process. An instance of the HIS-DF for the specification of integration between public health and clinical information systems has been previously discussed by the authors [4].

2. Formally Profiling HealthCare Standards

The paper describes a method and necessary tooling for identifying and creating UML profiles to health care standards, and applying them to existing UML models. The

method is integrated as guidance artifact into the HIS-DF Eclipse configuration. The method is summarized in the following subsections.

2.1. Identify Standard Models and Tasks in the Software Development Process

HIS-DF is the proposed framework for using profiles in an HIS development project. The task within the HIS-DF process is called standards harmonization. It can be defined within any of the three RM-ODP viewpoints described in the process. In the Enterprise Viewpoint, the task Business Use Cases can for example extend the HL7 profiles defined for the use case models, activity diagrams, business rules, storyboards, and other artifacts of the HL7 Development Framework (HDF). In the information viewpoint, profiles can be defined to formally describe the system analysis model based on HL7 RIM, DIMs, Refined Message Information Models (R-MIMs), Common Message Element Types (CMETs) or Templates. Finally in the computational viewpoint, profiles can be defined for HL7 services specifications, application roles, trigger events, interactions, and messages specifications.

2.2. Implement the Standard as a UML Profile

After having identified the models and the processes in the HIS-DF methodology where a standard can be reused, it is represented as a UML profile. The classes and properties of the information model of the standard are described as stereotypes and tagged values respectively. The Object Constraint Language (OCL) can be also used to constraint the profile. In this stage, UML tools are necessary to build the UML profile.

2.3. Apply the profile to System Models

The previously defined UML system models are annotated with the standard models by applying the profile. Conformance to the standard is assured because the UML profiles impose certain restrictions on how the UML model can be modified, excluding further changes to the profile. These restrictions should be supported by the UML tooling, however. A mapping (applying rules manually or tool-assisted) describes which classes from the original model corresponds to the base classes in the profile. Using model transformations it is possible to automate the mapping process.

2.4. An Example using HL7 Profiles

In order to exemplify how the previously described method is applied to an HIS development project, a public health information system example is used. A similar scenario has been described in a previous paper [4], but especially focusing on the use of the HIS-DF methodology. In that work, the information models were manually extended using HL7 Visio tools. That approach didn't follow the UML profiling mechanism, resulting in HL7 models incompatible with the UML 2.0 standard.

2.4.1. Identify Standard Models and Tasks in the Software Development Process

The method in this example supports the informational viewpoint described in the HIS-DF, concretely the system analysis model. The main diagram in the model is a UML class diagram describing the system information entities, its properties and relation-

ships. Figure 1 describes the analysis model for the public health surveillance system as a UML class diagram. The main classes in the domain are represented: Citizen, Event-Place, Public Health Authority and the Health Service Provider, all being associated to a Public Health Event.

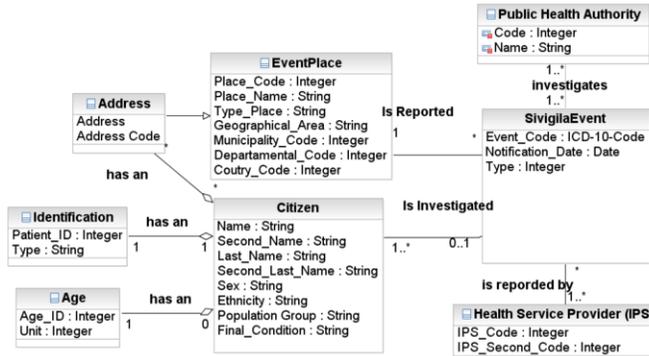


Figure 1. An Analysis Model for a Public Health Surveillance System

Following the harmonization process defined in the HIS-DF methodology, each one of the classes in Figure 1 are mapped to classes, attributes and data types defined in HL7 information models. Table 1 summarizes the mapping to classes in the HL7 RIM and successive refinement to classes in the Notifiable Condition Report R-MIM (PORR_RM100001UV01) as part of the HL7 v3 Normative Edition 2006.

2.4.2. Implement the Standard as a UML Profile

According to the mappings found in Table 1, UML profiles for the HL7 RIM and the Notifiable Condition Report R-MIM are necessary. The UML profile for the HL7 RIM is imported from the HyperModel tool [5], whereas the R-MIM profile is created using the IBM Rational Software Architect tool. The structure of the R-MIM profile is shown in the “Project Explorer” view in the left frame in Figure 2. A UML stereotype is created for each entity, act and role of the R-MIM, along with its attributes. The profile is made available for reuse as an Eclipse plug-in and an XMI file.

Table 1. Mapping the Surveillance System model to HL7 Information Models

Sivigila Classes	RIM Core-Classes	HL7 RIM Classes	R-MIM Classes
SivigilaEvent	Act	Public HealthCase	PublicHealthCaseEvent
Health Service Provider	Role	RoleHeir	AssignedEntity
Citizen	Entity	Person	InvestigatedPerson
EventPlace	Entity	Place	EventPlace
Public Health Authority	Role	RoleHeir	TerritorialAuthority
Address	Attribute	Person.addr	InvestigatedPerson.addr
Age	Attribute	Person.birthTime	InvestigatedPerson.birthTime

2.4.3. Applying the Profile to System Models

Figure 2 (main frame) shows the analysis model for the public health system, after having applied the HL7 profiles. For each class in the model, the applied stereotype is represented, along with the HL7 attributes and data types extended from the HL7 RIM and R-MIM information models. For example the SivigilaEvent class extends the <<PublicHealthCase>> and <<PublicHealthCaseEvent>> stereotypes defined in the RIM and

R-MIM profiles, respectively. The classes are colored using the HL7 functionality provided by the HyperModel tool.

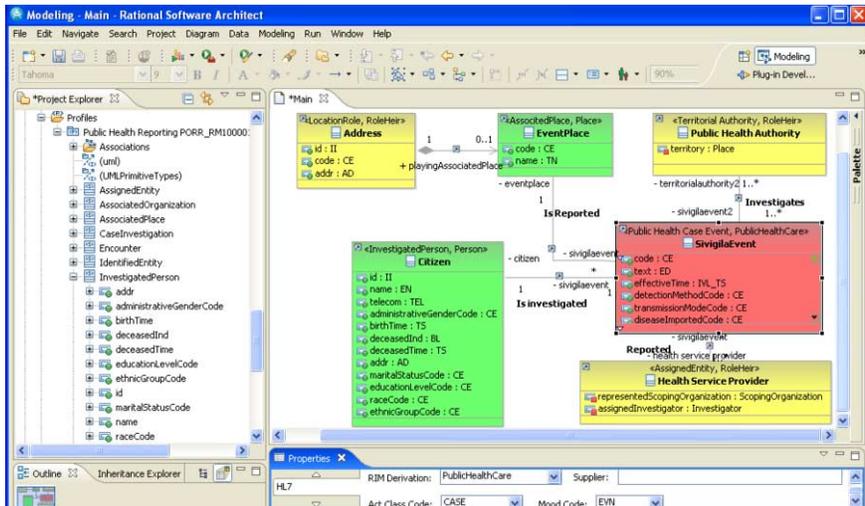


Figure 2. The Public Health System Information Model using the HL7 Profiles

3. Discussion

Few approaches suggest the use of UML profiles to support the development of semantically interoperable systems. In [6], a method for profiling healthcare standards using UML is proposed. The method is not integrated into a development process, and the creation of ad-hoc profiles is enforced. Carlson [7] developed a tool for MIF to UML transformation. While UML profiles for HL7 HDF and RIM have been developed, they have not been used for HIS modeling. CaCORE [8] provides an infrastructure to semantically annotate UML information models with specific domain standards and vocabulary, but not exploiting UML profiling mechanisms.

The proposed method for identifying and creating UML profiles to healthcare standards provides the following advantages:

- Support for Semantic Interoperability. System models are annotated, extending standard information models and intrinsic vocabulary. HL7 information models are at the moment the recommended set of standards to be reused because they cover a wide spectrum of clinical information, along with business processes, and services partially covering the business, informational and computational viewpoints defined in the HIS-DF methodology.
- Formal Mechanism for Models Specialization. An UML profile provides mechanisms for specializing healthcare standards in such a way that the specialized semantics does not contradict the semantics of the generalized standard. The profile establishes restrictions on how the standard should be specialized; avoiding changes in the meta-class definitions.
- Standard and Shareable Models. UML is the standard de-facto language for software systems modeling. UML models and profiles can be created and shared using the XMI standard. Most UML tools support the implementation of profiles.

- Support of Model-Driven Development. The UML system models described are platform independent information models. They can be transformed into platform specific models and code (Web Services, J2EE, and CORBA) using UML transformations. Profiles can also support transformations (mapping) between standards.

The major limitation of using existent UML modeling tools (even using the IBM Rational family) is that functionalities to specialize reference information models are not completely supported. Visio R-MIM Designer provides such functionality, but the resulting models can not be processed by conventional UML tools, however. Further limitation results from incompatibility between existing tools. For example, in the tools used for this study despite being Eclipse based tools, incompatibility between the latest versions of the HyperModel and the IBM Software Architect occurred due to incompatibilities between the Eclipse SDK distributions. It can be expected that with HL7's tooling migration towards the Eclipse platform, the R-MIM Designer will be extended by functionalities to process UML models and profiles.

4. Conclusion

Analysis and design of health information systems, especially the underlying business and informational models describing basic concepts, business and relation networks, have to be based on standards. The paper describes a method and necessary tooling for reusing standard healthcare information models, by that way supporting the development of semantically interoperable systems and components. The approach is based on the UML extensibility mechanisms by defining profiles. Using UML provides several advantages such as tooling support, graphical notation, exchangeability, extensibility, code generation, etc., deployable in the next generation of HL7 tools.

Acknowledgments

The authors are indebted to thank their partners from HL7, OMG and other organizations for collegial collaboration. This work was partially founded by the German Academic Exchange Service DAAD and the University of Cauca in Colombia.

References

- [1] ISO/DTR 20514 Health informatics - electronic health record: definition, scope and context, 2005.
- [2] Object Management Group (OMG), UML 2.1.1 Superstructure Specification, 2007.
- [3] López DM, Blobel B. A Development Framework for Semantically Interoperable Health Information Systems, Submitted to the International Journal of Medical Informatics (ISSN 1386-5056), 2007
- [4] Lopez DM, Blobel B. Connecting public health and clinical information systems by using a standardized methodology. *Medinfo 2007*; 12(Pt 1):132-6. (Best Student Paper Award).
- [5] <http://www.xmlmodeling.com/hypermodel/overview/>
- [6] Walderhaug S, Mikalsen M, Hartvigsen G, Stav E, Aagedal J. Improving systems interoperability with model-driven software development for healthcare. *Medinfo 2007*;12(Pt 1):122-6.
- [7] Carlson D, Singureanu I. Importing MIF to UML and Generating XDS. *Veteran's Administration Research Report*. January 2007.
- [8] Phillips J. et al. The caCORE Software Development Kit: Streamlining construction of interoperable biomedical information services, *BMC Med Inform Decis Mak* 6 (2006) p. 2.

Reconciling Data Structures in Health Information Systems for Research in a European Clinical Registry

Thomas H. MÜLLER

*Institute for Medical Informatics, Biometry and Epidemiology,
University of Munich, Germany*

Abstract. The use of both, retrospective and prospective data obtained from the health care process in biomedical research meets with problems related to the respective data structures used in different health information processing systems. This issue is explored in an exemplified clinical registry for antiviral resistance. Essentially, the data structure used in clinical systems or data capture systems conflicts with the requirements of many established methods of statistical analysis. To resolve this problem efficiently, a set of generic transformations is proposed.

Keywords. Health information systems, European projects, Data models, Statistical analysis

Introduction

Biomedical and health care research often relies on retrospective and prospective data obtained from the health care process rather than fully controlled clinical trials. While this is in part motivated by the high costs of clinical trials, many scientific questions cannot be addressed in trials for a plethora of different reasons, including ethical considerations and even practical issues. However, clinical trials and clinical routine constitute vastly different processes, resulting in both, different conceptual data structures and data modeling schemes. In addition, using the Internet for conducting clinical research projects has gained popularity because of the increased efficiency of communication [1,2,3,4]. Yet, this new technology has (re-)introduced an old paradigm, i.e. the hierarchical structure, because web documents are basically structured in this way.

Current health information processing systems are often based on relational databases or even simpler forms of tabulated data. Hence, the applied modeling schemes are often closely related to the entity-relationship paradigm at the conceptual level. The most common schemes are summarized in Figure 1. The standard object and attribute tabulation model (A), where each attribute of an object, e.g., a patient or trial subject, is explicitly included in the relational representation. There is at least an implicit assumption that most positions in the table are filled, i.e. that more or less all attributes are available for all objects. Because of this, it is also the form most

commonly used for statistical analysis. The Study Data Tabulation Model (SDTM) of the Clinical Data Interchange Standards Consortium (CDISC) constitutes an effort to standardise this approach[5]. While the standard tabulation model can also be used in a clinical research for capturing a concise and invariable data set, it is quite impractical for clinical registries and even more so for hospital information systems, because these systems must accommodate a large number of different possible attributes while each individual instance usually provides but a few of those attributes. Hence, the majority data capture systems employ the object-attribute-value or entry-attribute-value (EAV) scheme (Figure 1B).

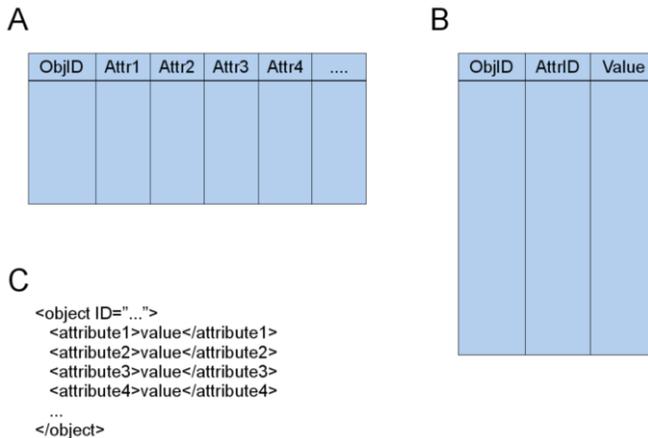


Figure 1: Common data modeling schemes in health information processing: Standard object tabulation with attributes (A), Entry-attribute-value (B), and hierarchical using XML (C)

The third scheme depicted in Figure 1C is not a relational one, but instead consists of a hierarchic structure. Specifically, this scheme is used in modern web technology via the standard extensible markup language XML and has also received a significant amount of attention in the standardisation efforts of HL7 [4]. It should be noted, however, that XML can also be restricted to resemble a tabular structure

This paper explores the data structure available from an example registry, the European registry for the surveillance of antiviral resistance in hepatitis, and proposes a methodological approach for reconciling those structures with the requirements of statistical analysis via a set of generic transformations.

1. Methods

The focus of this paper is on common relational and hierarchical data modeling techniques as they are used in many current health information systems. More conceptual or process-oriented approaches such as the Unified Modeling Language (UML) or the Zachman Framework are intentionally left out since the processes underlying the actual observations that ultimately comprise clinical registries are generally quite diverse and, furthermore, mostly out of the registries' scope or control. Likewise, object-oriented models aren't included since this technology is essentially absent in clinical data models. The data structures do, in fact, implicitly describe some

of the biomedical target processes of the research associated with the registries, but this aspect exceeds the scope of the present paper.

The web-based health information system developed for the European antiviral resistance surveillance registry (Project VIRGIL; see www.virgil-net.org) including the remote data entry as well as the data retrieval tools are entirely based on open source software components, including Linux, the Apache/Tomcat web-application server [6], the database management system PostgreSQL [7], standard programming languages (Java, Perl).

2. Results

2.1. The VIRGIL clinical registry

Figure 2 shows the main dispatch screen for entering case data into the European surveillance registry for antiviral resistance in hepatitis. This facility was started in early 2007. It is intended for any and all cases of antiviral resistance in (mostly chronic) hepatitis patients that are encountered by a participating centres across Europe. This means that patients are entered into this system at very different time points in their medical history. As a rule, these patients are subject to multiple cycles of antiviral therapy, with cycles usually lasting several months or even years. Resistance may occur as early as the first cycle or as late as decades after the first treatment.

The screenshot shows a web browser window with the URL <http://db.surveil.virgil-net.org/>. The page title is "VIRGIL SURVEIL Patient". The user is logged in as "Data Manager, System Administrator" (user thm). There is a "logout help" link. The main content is a data entry form with the following fields:

Center	D0003 MHH
Date of presentation	6/9/2006
PID	D0003:0003
Gender	male
Initials (optional, first and last name only)	
Year of Birth	1938
Country of origin	Germany
Diagnosis	Chronic Hepatitis B
Marital Status	married
Profession	
Occupation	unknown
Informed consent received ?	yes
Date of consent	6/9/2007

At the bottom of the form, there are several buttons: "Change", "Primary Documentation", "PriorTherapy", "FollowUp", "End Documentation", "Serum Samples", and "Home".

At the bottom of the browser window, there is a status bar with the text "Fertig".

Figure 2: European registry for antiviral resistance in hepatitis.

The registry accommodates this variability in patient history by not only providing data entry forms for follow-up visits (button "FollowUp"), but also for multiple episodes of prior diagnosis and therapy (button "PriorTherapy"). The entire data dictionary comprises about 1200 individual possible data items. However, only sparse

population of the full relation (outer join) is expected. The current proportion of filled attributes is about 17%, and is not expected to increase markedly.

A highly important property of this temporal structure is that it is necessarily uncontrolled and thus differs dramatically from the regular time scheme of visits usually imposed on controlled clinical trials. We must, therefore, extend the schemes from Figure 1 to accommodate the time axis along which the recorded observations take place, as . While there are obviously many possible ways to do this, an example of each that is deemed to be consistent with the underlying goal of the original scheme is shown in Figure 3.

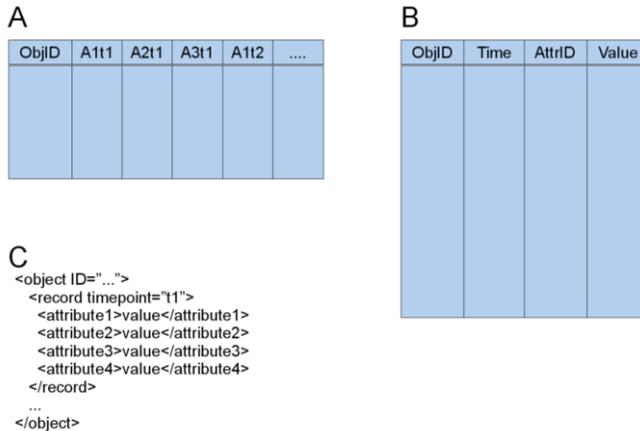


Figure 3: Modeling the temporal domain: Standard object tabulation with attributes(A), Entry-attribute-value(B), and hierarchical using XML(C)

The straightforward extension of the standard tabulation scheme is to replicate each attribute for each timepoint. This is clearly impossible to implement unless the timepoints are limited to a small grid. Even then, the resulting scheme is extremely “wide” and hard to manage. In order to make registry data available for statistical analysis in an efficient manner, a set of generic transformations is proposed.

2.2. Transformations

To begin with, it is clear that the EAV scheme and the hierarchical XML scheme shown in Figure 3 are equivalent as long as the depth of the hierarchical scheme is limited (non-recursive). One simply has to construct the attribute identifier space appropriately. The XML path language XPath would suffice. Next, the transformation from the standard table to EAV is also quite obvious. Finally, the interesting transformations lead from EAV, hierarchic XML, or sparse table to a full table. Such transformations may consist of the following :

1. Selecting the first, n^{th} or last value or date of a possibly recurring attribute,
2. Computing a statistic (e.g., minimum, maximum, mean, median,...) of all values of a recurring attribute,
3. Detecting an event (first, n^{th} , or last) such as a status change or a threshold crossing,

4. Detecting events and computing their frequency.

All of the above “elementary” transformations are non-invertible in general since they discard or aggregate data. The decision as to which transformation is appropriate must be made based on the research objectives in each instance. Table 1 shows a highly simplified (and biologically incomplete) example based on simulated registry data from patients with an assumed hepatitis C virus (HCV) infection. The following attributes are considered: an arbitrary patient identifier (PID), the blood test for alanine aminotransferase (ALT; in U/l) used to assess disease progression and therapeutic efficacy, the daily dose of the drug Ribavirin (in mg), and the viral particle count HCV_RNA (in IU/ml). The simulated data is similar to real registry content, but avoids any potential data protection issues.

EAV Data

PID	Date	Attribute	Value
1	22/11/2005	ALT	78
1	22/11/2005	Ribavirin	1200
1	22/11/2005	HCV_RNA	3400000
1	06/01/2006	ALT	48
1	16/03/2006	ALT	56
1	16/03/2006	HCV_RNA	8800000
1	22/05/2006	ALT	68
2	09/02/2006	ALT	60
2	09/02/2006	Ribavirin	800
2	09/02/2006	HCV_RNA	7130000

PID	Date	Attribute	Value
2	17/03/2006	ALT	58
2	17/05/2006	ALT	49
2	17/05/2006	HCV_RNA	6270000
2	29/05/2006	ALT	46
2	14/06/2006	ALT	25
2	14/06/2006	HCV_RNA	123000
2	12/07/2006	ALT	19
2	08/08/2006	ALT	14
2	08/08/2006	HCV_RNA	13700
2	06/09/2006	ALT	15

Transformation Result

PID	Ribavirin start date	Ribavirin	max. HCV_RNA	ALT normalisation	date of ALT normalisation	last HCV_RNA
1	22/11/2005	1200	8800000	no		8800000
2	09/02/2006	800	6130000	yes	12/07/2006	13700

Table 1: Example transformation from EAV (top) to standard tabular data (bottom)

By using the transformations listed above, a meaningful dataset that is amenable to classical statistical methods common in medical research can be derived as illustrated in the bottom part of Table 1. The purpose of the statistical analysis envisaged in this rudimentary example might be to assess the effectiveness of the antiviral drug. The Ribavirin start date and dose are derived via elementary transformation 1, assuming that the first therapeutic cycle is of interest. The maximum HCV_RNA level is an application of the second elementary transformation and could be used along with the last HCV_RNA in the assessment of drug effectiveness. ALT normalisation is the result of the event detection transformation based on a threshold of 20 U/ml.

3. Discussion

Managing data in clinical registries entails considerable problems related to the inherent structure of data collected during clinical routine, i.e., without adhering to a strict protocol. Furthermore, the attribute set recorded tends to grow over time as the corresponding medical field progresses. Whether one chooses classical flat (“horizontal”) tables or (“vertical”) EAV models to represent such data, the resulting relations are either broad and sparse or narrow and very long. Both cases are poorly

accommodated by database management systems and are ill suited for many statistical analysis methods established in the medical field. Corwin et al. [8] have developed solutions addressing architectural and performance issues related to sparse data. The proposed elementary transformations described above may provide a means to bridge the remaining gap since the transformation result is no longer sparse and can therefore be analysed using standard statistical methods. Obviously, these transformations can be adapted to any database architecture used to represent the sparse input data.

Viewed in a more abstract way, the proposed transformations essentially perform very basic time domain analysis functions on individual attributes. The results are then passed on as variables to the actual biostatistical analysis. From our experience in preparing registry excerpts for specific research projects, we know that even these basic functions are both, necessary and useful. There clearly are applications requiring other or more powerful transformations, such as integration/differentiation, multidimensional process analysis, frequency domain analysis or (auto-)correlation. Efforts of semantic integration such as the BRIDG Project[5] might at some point provide a means of automatically selecting an appropriate transformation and its parameters.

Due to the similarities, the proposed transformations may be equally useful in adapting data recorded during clinical routine for research projects.

Acknowledgement

This Project was supported by grants from the German Bundesministerium für Bildung und Forschung (Hep-Net, 01 KI 0403; Deutsche Leberstiftung) and by the European Commission (VIRGIL, Contract No. LSHM-CT-2004-503359)

References

- [1] Vincenzi B, Recchia G, Strozzi C, Scognamillo B: Clinical trial management and remote data entry on the internet. *Drug Inf. J* **33**, 1999.
- [2] Müller TH: Central IT-Structures for Integrated Medical Research and Health Care of Viral Hepatitis - Hep-Net. *Stud Health Technol and Inform.* **116**: 1016-20, 2005
- [3] Lopez DM, Blobel BG: Semantic interoperability between clinical and public health information systems for improving public health services. *Stud Health Technol Inform.* **127**:256-67, 2007.Klein A,
- [4] Ganslandt T, Brinkmann L, Spitzer M, Ueckert F, Prokosch HU: Experiences with an interoperable data acquisition platform for multi-centric research networks based on HL7 CDA. *AMIA Annu Symp Proc.* **986**, 2006.
- [5] Fridsma DB, Evans J, Hastak S, Mead CN: he BRIDG Project: A Technical Report.J Am Med Inform Assoc. 2007 Dec 20; [Epub ahead of print]
- [6] www.apache.org
- [7] www.postgresql.org
- [8] Corwin J, Silberschatz A, Miller PL, Marenco L: Dynamic Tables: An Architecture for Managing Evolving, HeterogeneousBiomedical Data in Relational Database Management Systems. *J Am Med Inform Assoc* **14**(1): 86–9, 2007

Using ESB and BPEL for evolving healthcare systems towards SOA

D. PAKONSTANTINOY, F. MALAMATENIOY and G. VASSILACOPOYLOS

Department of Digital Systems, University of Piraeus, Piraeus 185 34, Greece

Abstract. Healthcare organizations often face the challenge of integrating diverse and geographically disparate information technology systems to respond to changing requirements and to exploit the capabilities of modern technologies. Hence, systems evolution, through modification and extension of the existing information technology infrastructure, becomes a necessity. This paper takes a process perspective of healthcare delivery within and across organizational boundaries and presents a disciplined approach for evolving healthcare systems towards a service-oriented architecture using the enterprise system bus middleware technology for resolving integration issues and the business process execution language for supporting collaboration requirements.

Keywords. Evolution; healthcare systems; SOA; ESB; BPEL; standards; EMR.

Introduction

Healthcare organizations often invest significant resources in the development of large and complex information systems that must be modified and extended to respond to changing requirements, in addition to capitalizing on modern technologies [1]. Technologically innovative efforts of the past have often resulted in the development of disparate, incompatible and heterogeneous systems based on the traditional transaction processing paradigm rather than on supporting business processes. Thus, the integration of diverse and disparate information systems with emphasis on communication and collaboration is a challenge often faced by healthcare organizations [1,2]. To meet this challenge, a systems evolution process must be designed and implemented with the objective to achieve interoperability among diverse systems that may have been developed at different times and with different technologies, and to support a process view of the healthcare delivery context.

System evolution is an iterative and incremental process directed toward long-term user needs and operating on legacy or existing systems [3]. Having a certain direction, evolution differs from an unconstrained series of small modifications that may have little direction over long term. It also differs from systems development that has a specific direction or unifying intent but can start from scratch. Thus, the complexity and volatility of requirements for large-scale healthcare systems and the large in-place investments make evolution a necessity.

In most cases, the decision for system evolution hinges on whether the existing system architecture and functions fit current and anticipated requirements of the

problem domain. Thus, in evolving existing systems, most weight must be placed upon the need to explore the systems operating context, develop this context where necessary and make the target system serve it. When this context is described in terms of its constituent business processes and interactions among them, as is often the case, the evolution process focuses on designing a system in terms of how it supports business processes.

When an integrated, process-oriented healthcare system is envisaged, developers are required to first solve communication-level integration issues, ensuring that existing systems using different transport protocols and data formats can exchange information, and then to decide how existing systems can interact to support business processes [4]. Along these lines, this paper presents a healthcare systems evolution process that is directed towards the Service-Oriented Architecture (SOA) philosophy using the Enterprise Service Bus (ESB) middleware and the Business Process Execution Language (BPEL) to support systems integration and collaboration.

1. A system evolution process

The drive in healthcare to contain cost and improve quality has called for a transition from the institution-centered to consumer-centered care, requiring increased cooperation and collaboration among functional units. This creates an impetus for healthcare organizations to evolve their existing systems so that to enable integrated access to patient information irrespective of the location it resides. SOA, as a set of guidelines for integrating disparate systems, represents a view as to how business and technology architectures can be integrated and how composite application functionality is delivered to a portal or Web browser. Hence, SOA constitutes a promising evolution intent for healthcare systems since technological heterogeneity is more the rule than the exception in most healthcare organizations [1,2,3].

SOA is a concept that enables end-to-end application integration across and among healthcare providers, and provides a flexible model that permits the healthcare organizations to respond to environmental changes quickly and efficiently. It constructs applications from the ground up, taking one or more services and connecting them to form a complete, cross-functional, end-to-end healthcare process such as "medical order". This architectural style is dependent on the Internet and is a combination of business architecture, application architecture and software architecture. In addition, it allows users to rapidly build, reuse and reconfigure automated workflow processes (services) as healthcare priorities, regulatory requirements or environmental conditions change [5]. Hence, SOA responds to the need of exchanging medical information between diverse healthcare systems on the web and can form an ideal architectural basis for evolving existing healthcare systems [6,7,8].

A SOA environment needs a robust and secure infrastructure that can easily combine and re-assemble services to meet changing requirements without disruption and can span new service-enabled applications as well as existing [9]. ESB meets the connectivity needs of applications and services, by matching and routing messages between them, and makes services available for broad access and reuse [9,10]. An ESB operates as a bus that connects the various resources, ensures that services are exposed over standards-based protocols (e.g. SOAP, HTTP and JMS), enabling any client to contact them directly and to perform transformation and routing of service requests. ESB increases system performance, thus enabling the development of virtual healthcare

applications as a connected, process-based set of independent services that exist inside or outside the healthcare organization. Thus, ESB facilitates cross-platform interoperability, unifies message oriented, event driven and service-oriented approaches for integrating applications and services, and, hence, provides a technological basis for evolving existing healthcare systems into an integrated district-wide environment [9,10].

In a federated healthcare environment (i.e. one in which services cross organizational boundaries) or in a distributed healthcare environment (i.e. one in which service communications bridge geographic boundaries) it is critical that data, events and replies are directed to the right place at the right time without management overhead. To this end, an ESB, besides resolving integration issues, provides the capability of orchestrating healthcare activities into processes, typically via BPEL [4]. BPEL provides a standard, XML-based platform that expresses a business process's event sequence and collaboration logic, whereas the underlying services provide the process functionality. Hence, in a SOA approach that aims at supporting healthcare processes, BPEL can be used to implement intra and inter-process collaboration.

Based on the above, a system evolution process towards a process-based SOA consists of the following stages:

- i. Develop a process description of the healthcare organization, or part of the organization, under review and specify process support requirements, including collaboration and cooperation requirements. To this end, generate a graphical representation of how and in what order process activities are executed as part of a human workflow.
- ii. Map existing system applications onto the process models developed to compare "what is available" with "what is needed". From this comparison, identify which applications to keep, which to modify or upgrade, which to create or develop and which to discontinue so that the target system meets the requirements. Then, develop old and new applications as services that are assigned to process activities.
- iii. Describe services into the BPEL engine, that generates an XML file which tells the message router (service bus) the rules associated with each service, and use ESB as a message router and rules engine that performs validations on messages, transformations on data and ensures message delivery from one system to the next, and eventually the delivery of information to the user's portal or browser. Thus, a message is simply the packet of data requested by any service.

The above approach suggests an evolution direction where a SOA implemented on an ESB/BPEL software infrastructure becomes a vital issue in supporting relationships such as those which exist between the healthcare systems and their operating context, relationships between roles at work and relationships between parts of the systems which are expressed as services. This paper focuses on the third stage.

2. Motivating scenario

To illustrate the technical aspects of the above approach to systems evolution, a sample integration project is described which is concerned with the automation of cross-organizational healthcare processes spanning a health district. Typically, a health district consists of one district general hospital and a number of peripheral hospitals

and health centers. As patient referrals are usually made among various healthcare providers within a district, there is a need to ensure that access to integrated patient information by authorized users is enabled when and where needed. Thus, the sample process considered here is concerned with patient referrals from health centers to hospitals and it involves three separate systems based on different technologies:

- *Radiology Order System (ROS)*: a system that handles medical order processing among healthcare organizations and is already exposed as a web service.
- *Radiology Report System (RRS)*: a system that uses a Java Message Service (JMS) queuing system for communication.
- *Electronic Patient Record (EMR)*: a customized system implemented in Corba.

Suppose a healthcare process that begins with a health center's physician request for a radiological procedure on one of his/her patients and ends with issuing a radiological report by a radiologist. In this process two functional units are involved: the health center and the radiology department of a hospital. On performing the radiological procedure requested, the radiologist accesses the relevant part of the patient record and issues a radiological report, incorporating both the radiological images and the associated assessment, which is sent to the requesting physician. Figure 1 shows a high-level view of the healthcare process concerned with radiology orders.

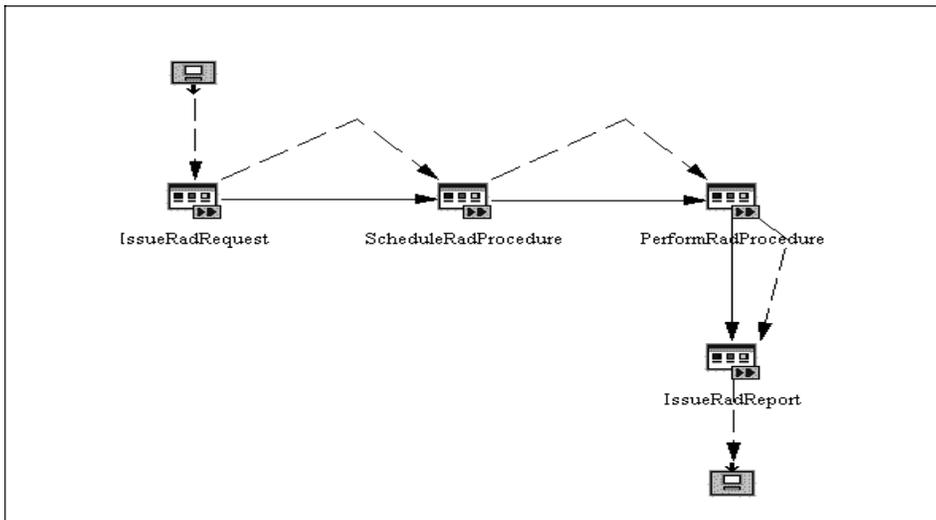


Figure 1. A high-level view of the healthcare process concerned with radiology orders

3. System architecture

The system architecture described here delineates the intent of an evolution process directed toward a SOA implementation on an ESB/BPEL software infrastructure to enable access to integrated patient information during the execution of healthcare processes spanning a health district. To this end, the sample applications mentioned in the previous section are synthesized within a SOA so that they serve as a unified whole.

Uniting these applications into business processes involves solving various integration issues by exposing each application as a web service and, then, using BPEL

to combine the services into business processes. Figure 2 shows a schematic view of the architecture's four main components. These are: the existing IT infrastructure; the ESB which includes adapters to expose existing systems and provide transport connectivity; the BPEL engine that is capable of interpreting and executing business processes described in BPEL by orchestrating existing services and the services developed or created from the existing systems. A fifth component may also be introduced to represent an authorization server that manages who, in terms of role, can perform the various BPEL activities, such as invoking a service or assigning a new value in an XML document.

In the context of the proposed architecture, the ROS is already implemented as a web service and, hence, no further development is required. Additional work is only required for exposing the two remaining systems as services. Specifically, an ESB allows clients to access the services through HTTP (or other protocols) and forwards client requests to the RRS via JMS. In this context, new message formats are defined using a CDA-based XML Schema and transformation rules are created to convert to the existing application's format. This results in a new ESB-based web service for RRS, which receives requests and transforms them before placing them in the JMS queue and communicates with other web services via BPEL. Figure 3 shows the BPEL constructs that were added to WSDL files for this purpose. The PartnerLinkType defines the role of each partner in a healthcare process and ties it to a given *PortType* (i.e. the WSDL term for interfaces). In this example, the presence of a single role implies a client-server relationship. The figure also shows the *RadProcCd* property definition and a property alias that describes how to extract that property's value from the authentication message.

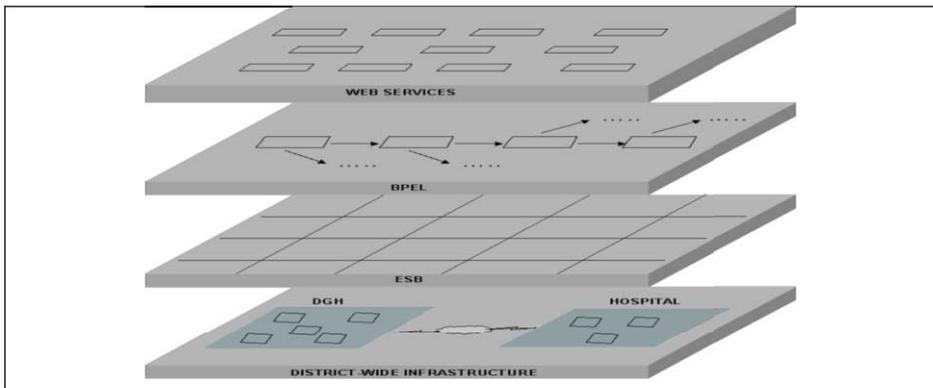


Figure 2. A service-oriented architecture implemented on ESB and BPEL

```

<plnk:PartnerLinkType name="RRSLinkType">
  <plnk:role name="serviceProvider">
    <plnk:PortType name="tns:RRS" />
  </plnk:role>
</plnk:partnerLinkType>
<bpws:property name="RadProcCd" type="xsd:string" />
<bpws:propertyAlias messageType="tns:authenticate" part="request" propertyName="tns:RadProcCd"
  query="/xsd1:EMRDetails/RadProcCd" />

```

Figure 3. Example of BPEL's WSDL extensions.

Finally, the Corba-based EMR must be addressed and to this end an ESB wizard can be used to automatically create the web service by designing the interface in WSDL and create the web service from there. The service implementation acts as a client to the Corba system directly or through an ESB-generated web service interface.

This architecture is compatible with the EN 12967 “Health Informatics Service Architecture” (HISA) standard, which aims at providing a reference model for health care IT services, facilitating the building and purchasing of interoperable systems [11]. Thus, healthcare information is clearly separated from the applications and can be made available where and when needed to the various modules of the information system.

4. Concluding remarks

Healthcare organizations are faced with the challenge to improve healthcare quality, preventing medical errors, reducing healthcare costs, improving administrative efficiencies, reducing paper work and increasing access to affordable healthcare. To these ends, innovative health information technologies are often used in order to shift focus from traditional transaction processing into communication and collaboration. This paper describes an approach to evolving existing systems towards a certain direction, namely, a SOA-based solution which is based on the ESB middleware technology to resolve integration issues and BPEL to orchestrate individual healthcare activities into healthcare processes. Thus, through the evolution process presented in this paper, communication-level integration issues are solved first to ensure that all partial systems (existing and new) using different transport protocols and data formats can exchange information and, once these issues are resolved, the various IT systems are made to interact to support healthcare processes. Hence, the result of evolution is an interoperable process and service oriented healthcare information system.

References

1. Lenz R, Kuhn KA. Towards a continuous evolution and adaptation of information systems in healthcare. *Intl J of Med Inf* 2004; 73: 75-89.
2. Huizen Van G and Backman AJ. SOA: First Principles. SONIC software, 2005.
3. Ferrara FM. The CEN healthcare information systems architecture standard and the DHE middleware – A practical support to the integration and evolution of healthcare systems. *Intl J of Med Inf* 1998; 48:173-182.
4. Pasley J. How BPEL and SOA are changing web services development. *IEEE Int Comp* 2005; 60-67.
5. Erl T. Service-Oriented Architecture (SOA): Concepts, Technology, and Design. The Prentice Hall Service-Oriented Computing Series from Thomas Erl 2005.
6. Bloomberg J. Service-Oriented Architecture: why and how?. Zapthink Whitepaper, 2003.
7. Fisher M and Jendrock E. Chapter 1: Introduction to web services. *Java Web Services Tutorial*. Sun Microsystems, 2002.
8. Schulte R. SOA is changing software. Gartner Research, 2002.
9. Davidsen L. Building an ESB without limits. IBM Software Workgroup, 2007.
10. Balani N. Model and build ESB SOA frameworks - Adapt service-oriented architectures for easy application integration. IBM DeveloperWorks, 2005.
11. Bott OJ. European Standards for the Electronic Health Record. eHealth benchmarking 2006. http://www.ehealth-benchmarking.org/2006/images/stories/10_bott_european_standards.pdf

Systematizing medical alerts

Mattias PETTERSSON^a, Jenny WIHLBORG^a, Rikard LÖVSTRÖM^{a,b,c},
Erik SUNDVALL^{a,1}, Mikael NYSTRÖM^a, Daniel KARLSSON^a
^a *Department of Biomedical Engineering, Linköpings universitet, Sweden*
^b *Boda Primary Health Center, Västra Götalandsregionen, Sweden*
^c *Humanhälsan i Väst AB, Sweden*

Abstract. The current Swedish regulations for medical alerts in health records were designed for paper records. Suggestions for computerized systems are now being investigated. A proposed model using three alert categories, graphically represented using three directions, probably combined with three severity levels is presented here. Up represents hypersensitivities, left/back represents alerting diagnosis and right/forward represents alerting current and planned treatments. A small qualitative user study of the alert classification model and some graphical representations of it was conducted. One main finding is that most respondents found the use of directions intuitive as a means of presenting categories. Context dependency, information overload, and future possibilities for automated alert-gathering are also discussed in the paper.

Keywords. Medical Warnings, EPR-CPR-EMR, Patient Safety, User-computer interface, Data acquisition- data capture, Graphical Representation

Background

To avoid exposure to known risk factors it is essential that important medical facts are put front and center. Lack of such information is a direct cause of injuries and even deaths in today's health care [1]. Current directives from Sweden's National Board of Health and Welfare require that certain important facts are presented highly visible [2]. Due to the fact that the regulation in particular is from the early 1980's it is intended for paper-based health records. It says that information about hypersensitivity should be indicated on, for instance, the cover of the record's folder with a red stamp or sticker.

During the last decades computers have been introduced as a tool in health record management. Trying to comply with legal standards suppliers of systems for electronic health records (EHRs) have implemented functionality in analogy with the cover stamps. Most common are icons using exclamation marks and triangles, often in colors like red, yellow and black [3]. The underlying data is still primarily related to drug hypersensitivity but may also give an opportunity to add other information that colleagues should know about. Research on alert indicators is hard to find on the international arena. Some parts of the work on VCM pictograms by Lamy et al [4] concern this area and may be applicable.

¹ Corresponding author: Erik Sundvall, Linköping University, Dept. of Biomedical Engineering, SE-581 85 Linköping, Sweden. erisu@imt.liu.se, <http://www.imt.liu.se/~erisu/>

In early 2007 a series of reports presented an information model for an “awareness signal” [5, 6]. The model was based on a broader definition than what is used in the present regulation. Among other things they mentioned current treatments and patient preferences, like declining to receive blood products and similar procedures.

Besides the extended definition a three-level scale of severity was introduced [5]. The most severe level is “life-threatening”. One step down is “harmful”, a condition bad enough to cause permanent and irreversible damage. The third one is “discomforting”, a non-lasting state that may lead to momentary or temporary issues. These levels are similar to the three used in some archetypes from openEHR², for instance the evaluation “Adverse reaction” [7]. In the openEHR setting the options for reaction severity are labeled “life-threatening”, “disabling” and “mild” respectively.

1. Material and methods

This paper seeks to present an initial visualization for an expanded definition of medical alerts and provide a first reaction from health care staffers. The main purpose has been to find a new standardized set of parameters for an awareness signal in the area of medicine. To narrow down the scope Swedish conditions have been our focus.

1.1. Expanded alert definition

Introducing an expanded set of parameters for medical alerts requires serious consideration. First of all the expansion must fit the clinicians’ needs. Secondly the parameters must be possible to clinically define and store in and retrieve from EHR systems. Finally there must be a way to present the data in relevant views to other clinicians. All of these criteria must be met in order to reach the fundamental requirement for a medical alert: the ability to show information that will prevent possible patient injuries; injuries that would occur if the alert was missing and standard procedures and treatments, harmless for most other patients, were used.

Starting with Carelink’s suggestions [5] for new alert parameters only a few made the final cut: hypersensitivity, current treatments and diagnoses. The use of the first one is already regulated and the two latter is added to widen the spectrum of information. Together these three include areas that need to be covered for more nuanced evaluations of the patient’s condition and treatment planning. Treatments can include everything from medications, which frequently give interactions, to implants—like pacemakers or surgical clips and staples—and similar prosthetics which are important to know about prior to some examinations, e.g. MRI scans. Diagnoses could embrace diseases resulting in a number of hypersensitivities, like malignant hyperthermia.

1.2. Visualization

In the prototype³, shown below, the broadened set of alert parameters is encoded in a graphic representation. This symbol includes both the extended information and the

² <http://www.openehr.org>

³ Developed as part of a master thesis project at Linköping University, LiU-IMT/MI20-EX--07/457--SE (to be published spring of 2008).

severity levels and by using simple associations all new parameters is assigned a direction. The names of the parameters correspond to the directions according to:

- **Up: hypersensitivity**
The Latin prefix *hyper-* translates to “over” or “above” as well as emphasizing increasing activity.
- **Left/backwards: important diagnoses**
Clinical findings are made now or have been discovered earlier. Seeing the lower axis as a timeline, with the present at the center, puts diagnoses on the left side.
- **Right/forward: treatments**
Current and planned treatments and other upcoming activities are set from now and into the future. Using the timeline metaphor these are put on the right side.

To indicate the severity of potential complications each direction is divided into segments to fit the three-leveled model. A short axis is equal to *discomforting* and a long is equal to *life-threatening*. This way the image seems larger the more severe the conditions are. The setup using perpendicular directions is meant to make complex information available in a structured and straightforward way. It shall give the user a first glimpse and an opportunity to later dig deeper into relevant areas. Even from a distance of a few meters away from the computer screen it gives a first impression of the situation. The colors are chosen to attract attention in contrast to the standard gray.

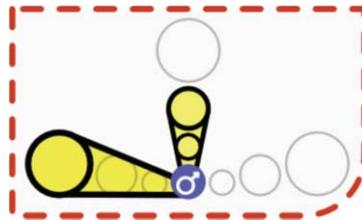


Figure 1: A sample symbol with three active alerts: one regarding a diagnosis, which due to some medications could cause a life-threatening situation (e.g. acute intermittent porphyria), one harmful hypersensitivity (e.g. gluten to coeliac disease) and one discomforting hypersensitivity (e.g. antibiotic causing significant diarrhea). The red border emphasizes a life-threatening warning and the center is used to indicate the patient’s sex.

Development and early evolution of the visualization were made with Bertin’s visual variables [8, section 4.3.1] in mind. Planar dimensions, size and colors are therefore the basic elements in the transition from structured data to visual object. Most of it is static, only colors lighting up the shapes when changing between the states. Using a grey outline for the inactive parts make it stay in place as well as, if empty, indicating that the alert system is implemented and running.

In order to get information about the single alerts there must be links between the symbol and the EHR notes. Using tool-tip functionality, like in many computer settings today, a short summary can be presented when the relevant area is hovered with the cursor. Taking yet another step the complete underlying note is retrieved.

It is important to point out that the alert does not have to indicate only incidents that already have happened. Issues that may occur also have an opportunity to be raised. The idea is to present the evaluation of future risks. Another crucial fact is that a gray symbol under no circumstances should be interpreted as if the patient can take any

treatment. Unknown or unrecorded problems may surface at any time. Empty symbols call for collection of information from other sources, perhaps from the patient herself.

1.3. Interviews

To make sure that a clinical perspective was present at all times, physicians have been consulted, interviewed and involved during all phases of development. For external feedback a short qualitative study involving six clinicians from diverging medical specialties was carried out. The interviews lasted 20-30 minutes and started with a short description of the project's initial outcome including the graphical representation. After the brief introduction it continued with questions related to their present use of alerts. In the end they had an option to propose suggestions for further work. The interviewed physicians were working in areas from general practice and anesthesiology to emergency medicine, everyone with their own perspective on the subject.

After all interviews the material collected during the conversations was processed by two of the authors, who also performed the interviews.

2. Results

Due to the wide range of use cases in the health care environment a large study is needed to capture thoughts and ideas on a subject like this one. Because of the nature of this study—exploring a newly introduced concept—this initial inquiry was set as a small-scale study. For wider conclusions further interviews needs to be performed⁴.

2.1. Discussing current use

Today's stickers and stamps for alerts—on or off, no nuances—sometimes give misleading information according to the respondents. A sticker, which should indicate something important about the patient's health, occasionally hints about personal preferences. On this the saying "Don't cry wolf too often" repeatedly came up in the interviews, demonstrating the importance that the information is correct and up to date. You must be able to trust it; otherwise the alerts are disregarded in the long run.

2.2. Introducing the new ideas

Most of the respondents got the main idea, regarding separated categories and their assigned directions, from the introductory explanation. The associations made them aware of where the different parameters were positioned, which made some of them able to intuitively refer back to the directions later in the interview. On the other hand everybody did not find use for separating the information into different categories. One respondent, who had a background only using paper-based records, wanted a simple symbol to indicate potential alerts; a symbol which either was on or off.

The severity assessment was found to be a part many had opinions about. A couple of the respondents believed they would benefit from the risk levels and that it was easy

⁴ Since this paper was first submitted, Nov. 18 2007, Carelink has published a report covering a more in-depth analysis regarding alerts. Available from: http://carelink.se/dokument/tillgang_till_vardinformation/nationell_patientoversikt/Rapport_Vaming2_080208_ver_1.0.pdf

to estimate them, while others did not see the distinctions. Some concerns also arose regarding increased workload. Assessing and tagging risks was expected to increase administration, thus giving less time for patient care.

2.3. Suggestions from the respondents

A range of suggestions for improvements and additions came up. The emergency staffer wanted an indicator for do not resuscitate orders. On the other hand the primary health physician wanted to know about reduced cognition and need for an interpreter. Also pregnancy came up as a suggestion. Some of the anesthesiologist proposed a more detailed graphical representation including the organ systems at risk.

Regarding display of patients' personal preferences most of the respondents, upon being asked, found it improper. The situation may change the mind of the patient and friends and family are usually rather quick to notify the care provider about preferences. Acting on old subjective information in an EHR may lead to wrong conclusions and an undesired outcome. The most common proposal was to add a warning for contagious blood, a wish relating to the staffs own health and work environment.

A risk mentioned was information overload. Since some alerts only concern a small set of specialists—hypersensitivity to some anesthetic agents are only relevant to anesthesiologists—it might clutter the visualization for others to a point when facts are lost. A solution suggested was context control, so that irrelevant information is hidden.

Finally it was pointed out that the EHR is only one of many tools. A dialogue between the patient and his or her physician is still an important source of information.

3. Discussion

The alert symbol shows a simplified version of the EHR content. It is a projection of relevant facts onto a pre-defined image outline. Making the most important record data easily available to the clinicians severe damages and deaths are hopefully prevented.

A new set of alerts categories must be comprehensive but minimalistic. In the same time as many areas should be covered—seen from the suggestions the clinicians made in the interviews—it must be kept to areas that more than a few medical specialties are interested in and helped by. The latter is important because if you do not feel use for the work you put in, you are less likely to do it well. And medical alerts must be done right.

3.1. Interviews

From the interviews it can be said that directions as information carriers is an area to keep investigating. How the respondents apprehended and learned the introduced concept from only a short presentation was notable. Using the associations from the categories' names made the interpretation easy and in some cases intuitive, from the look of it. The important thing is that the symbol can be interpreted from a distance and give guidance in how to proceed; either to continue as always or stop to take in the new facts. Some of the proposals made by the respondents had good merit. The suggestion regarding contagious blood is important for everyone in touch with a health care establishment, staff or not. It is essential not spreading disease, especially when it comes to blood. In some form this should be taken into consideration in future work.

Concerning the doubts on severity levels this needs further study. For whom is it important and for which parameters does it add value?

3.2. Automated alert gathering

A risk mentioned was increasing workload relating to assessing and adding new alerts. This could be addressed by computerized support. On a day-to-day basis diagnoses and treatments are documented, making most of the parameters already present in the EHR. Hypersensitivity is often related to chemical substances featured in the ATC⁵ classification and diagnoses are labeled with ICD-10 codes. Using these as base there are simple links between the three directions in the graphical representation and some easy to identify entities. Regarding treatments a combination of ATC and other classifications are suitable.

A mapping connecting the classifications to certain alert levels is one way to ease the burden. There are for instance lists of rare diseases [9] suitable for important alerts. How these pre-set alert levels are set is an issue for qualified medical personnel to sort out. Crucial to the automatic alert is to avoid false positives. Information given must be trustworthy. To assert the quality, the final assessment must be made by a human.

3.3. Closing comments

To conclude things we have found a demand and a wish for new alert categories, expanding the definition that today only include hypersensitivity. In addition the response to the visualization graphics, using directions to carry the data, felt positive and worth further exploring. This has already begun in a project by Carelink.

From the study we also found that in addition to alerts for the safety of the patients there is a need to alert the staff about risks they are exposed to, like contagious blood.

4. References

- [1] Socialstyrelsen [homepage on the Internet]. Om vårdskador. [updated 2007 November 12; cited 2007 Nov 15]. Available from: http://www.socialstyrelsen.se/Amnesord/patient_klientsakerhet/specnavigation/Kunskap/Vardskador/
- [2] Socialstyrelsen. Socialstyrelsens föreskrifter om markering i journal av intolerans och allvarlig överkänslighet mot läkemedel m.m. [regulation]. SOSFS 1982:8.
- [3] Wisser C. Användbarhet på dödligt allvar! Master thesis 2004:17. IT University of Göteborg, Department of Computing Science.
- [4] Lamy, J.-B., Duclos, C., Rialle, V., Venot, A. Which graphical approaches should be used to represent medical knowledge?. *Studies in health technology and informatics*, 116:719-724, 2005.
- [5] Carelink. Nationell patientöversikt 2006, varningsinformation. 2007.
- [6] Carelink. RIV Informationsspecifikation, verksamhetsdokumentation för varningsinformation. 2007.
- [7] openEHR[archetype on the Internet]. Adverse reaction: Evaluation [updated 2007 April 13; cited 2007 October 27]. Available from: <http://svn.openehr.org/knowledge/archetypes/dev/html/en/openEHR-EHR-EVALUATION.adverse.v1.html>
- [8] Andrienko N, Andrienko G. *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Berlin: Springer; 2006.
- [9] Socialstyrelsen [homepage on the Internet]. Socialstyrelsens kunskapsdatabas om ovanliga diagnoser. [updated 2007 November 12; cited 2007 November 16] Available from: <http://www.socialstyrelsen.se/ovanligadiagnoser>

⁵ <http://www.whocc.no/atcddd/>

A Framework for Semantic Interoperability in Healthcare: A Service Oriented Architecture based on Health Informatics Standards

Amanda RYAN ^{a,1} and Peter EKLUND ^b

^a *Faculty of Commerce, University of Wollongong, Australia*

^b *Centre for Health Services Development, University of Wollongong, Australia*

Abstract. Healthcare information is composed of many types of varying and heterogeneous data. Semantic interoperability in healthcare is especially important when all these different types of data need to interact. Presented in this paper is a solution to interoperability in healthcare based on a standards-based middleware software architecture used in enterprise solutions. This architecture has been translated into the healthcare domain using a messaging and modeling standard which upholds the ideals of the Semantic Web (HL7 V3) combined with a well-known standard terminology of clinical terms (SNOMED CT).

Keywords. Distributed systems, Modeling, Standards, Terminology-vocabulary

Introduction

Each year avoidable deaths and injuries occur because of poor communication between healthcare practitioners [1]. The *Systemized Nomenclature of Medicine - Clinical Terms* (SNOMED CT) is a universal health care terminology, the aim of which is “to improve patient care through the development of systems to accurately record health care encounters” [1]. Terminology is an important tool toward achieving semantic interoperability and thus improving communication in healthcare, and ultimately benefiting patients.

Terminology on its own only provides a standardised set of terms, but this model is strengthened by providing a standard structure for information and rules about where terms may be used in this structure. To this end, the *Health Level 7* (HL7) V3 messaging and modeling standard is used. HL7 is chosen because it is consistent with the Semantic Web vision of providing a universal and computer-interpretable medium for the exchange of data (specialised within the health care domain) [2].

1 Corresponding Author: Amanda Ryan, PO Box U124, University of Wollongong, Wollongong, NSW, 2500, Australia; E-mail: ajr883@uow.edu.au.

In previous work [3], a strategy for achieving semantic interoperability in healthcare by basing HL7 message models on SNOMED CT concepts and relationships was outlined. It was demonstrated that this methodology could be applied to a basic set of clinical observations successfully. Since then, a full PDA application has been developed for collecting clinical observations data which utilises these message models for communication with a PC. This work was done as part of the ePOC project, which has now concluded.

In this paper, the authors go on to describe an expansion of the basic message model construction method used in the ePOC project, and to outline their overall goal of creating a middleware infrastructure called the Health Service Bus (HSB), a service oriented architecture for effective communication in healthcare.

1. Methods

1.1. Health Service Bus

Enterprise Service Bus (ESB) is a term used to describe a middleware software architecture with a standards-based messaging engine, which is event-driven and provides foundational services for more complex software systems [4].

ESB is both operating system and programming language independent and provides interoperability between different platforms; for example, between Java and .NET applications. ESB is not necessarily Web-Services based (although this is often the case) and uses XML as a standard communications language.

A middleware infrastructure for health systems has been based on the ESB ideology, called the Health Service Bus (HSB). An overview diagram of the HSB architecture is shown in Figure 1. The HSB provides communication between disparate health systems which can all connect to it regardless of the type of software or hardware used. The HSB is also envisaged to provide services such as a terminology server and translation services for legacy systems.

In [3] a small subset of SNOMED CT *Observable Entities* with corresponding *Procedures* was concentrated on, in accordance to the clinical requirements of the ePOC project. Since then the focus of this research has been extended as demonstrator of the Health Service Bus. This involved taking a workable-sized subset of SNOMED CT² revolving around 'Vital Signs' terms and concepts, converting the subset to XML and then storing them in a native XML database.

The HSB is built on Jini technology, a service oriented architecture for the construction of secure, scalable distributed systems [5]. A Jini service for querying the SNOMED CT XML database has been created and is one of the core services of the HSB. Applications attached to the HSB communicate via JavaSpaces (part of the Jini Technology), using HL7 V3 in the actual content of the messages, in the format described later in this paper.

Figure 1 shows the ultimate goal of the HSB: the goal of PCs and mobile computing devices all accessing the HSB in order to communicate and access services such as terminology servers and patient databases.

² A subset is used because the entire SNOMED CT terminology is extremely large and not practical for research and testing purposes.

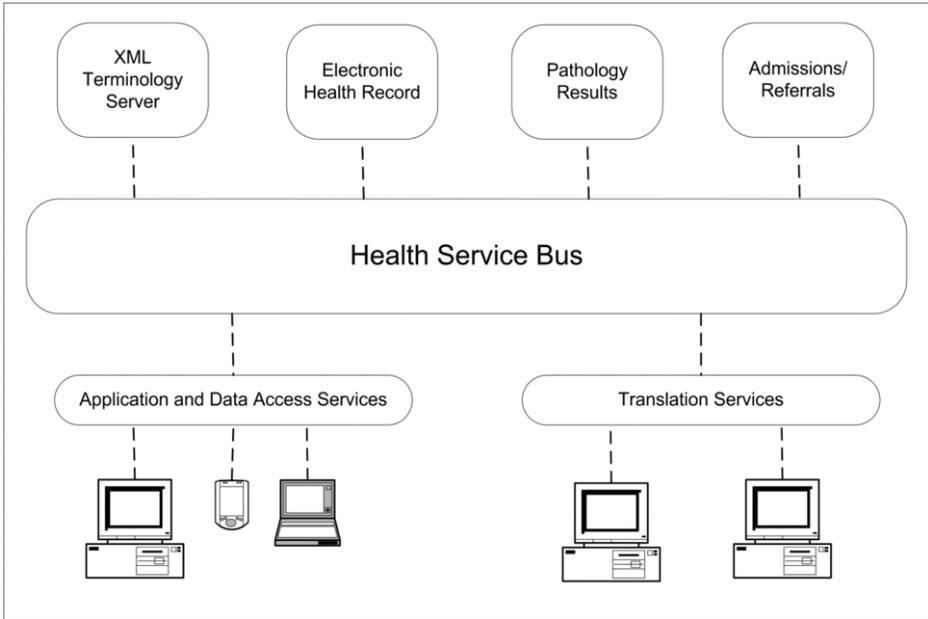


Figure 1. The Health Service Bus (HSB) architecture.

1.2. Previous Work

SNOMED CT concepts corresponding to a set of clinical observations were aligned with HL7 message models for use in the *ePOC* (*electronic Point of Care*) project. The *ePOC* project was a multi-phase, iterative R&D project with a research focus involving the development of a prototype PDA based *point-of-care* information delivery system [6]. A PDA application for clinical observations was developed for this purpose and these message models used and tested within this application [3].

The alignment of these message models is based on mapping SNOMED CT *Observable Entities* to HL7 *Observations* (a specialisation of the *Act* class) by populating the `code` field of the HL7 *Observation* with the code corresponding to the SNOMED CT *Observable Entity* concept. The relationships between the SNOMED CT concepts are upheld in the HL7 model by mapping them to *ActRelationships* between the *Observations*. At the same time, the HL7 `methodCode` field was populated with a SNOMED CT concept from the *Procedures* subset and the relationships between these concepts also upheld with the same *ActRelationships*. Table 1 shows a summary of the mapping between HL7 and SNOMED CT.

Table 1. Mappings of concepts and relationships between HL7 and SNOMED CT in ePOC.

Concepts	
HL7 Observation Field	SNOMED CT Subset
code	Observable Entities
methodCode	Procedures
Relationships	
HL7	SNOMED CT
component	Subtype
componentOf	Is a

1.3. Extended Alignment Strategy

The work described in the previous section (2.1), while being sufficient for its application, is a preliminary step toward a greater goal to create a complete framework for health communication and has formed the basis for the messaging models and use of terminology in this framework. To this end, the messaging focus has been expanded more generally to include all types of SNOMED CT concepts (not just *Observable Entities* and *Procedures*) by taking a subset of SNOMED CT based on the concept of ‘Vital Signs’. In this new strategy, HL7 models and messages are automatically generated based around the HL7 *PatientCareProvisionEvent* and *ObservationEvent* specialisations of *Act* using these terminological concepts.

Along with the mappings used in the ePOC project shown in Table 1, SNOMED CT *Body Structures* have been mapped to the HL7 *targetSiteCode* field, and *Clinical Findings* mapped to the HL7 *interpretationCode* field. This field is automatically populated based on the *code* combined with the *value* field, to assist with decision support.

As an example, if the *code* is *blood pressure* and the *value* is ‘120/80 mmHg’, the decision support system will automatically populate the *interpretationCode* field with the concept *normal blood pressure*, as the *normal blood pressure* concept has the relationship ‘interprets’ with *blood pressure*, and the blood pressure value is in a healthy range. The *targetSiteCode* field is then restricted to the *systemic arterial structure* and its subconcepts, based on the fact that the *normal blood pressure* concept has the relationship ‘finding site’ with *systemic arterial structure*.

Business rules (or constraints) have also been implemented to uphold restrictions between these fields so that all code values in the one instance of a HL7 *Observation* class must have existing relationships in SNOMED CT, or must not have contradicting relationships in SNOMED CT. These rules facilitate automatic message generation in the HSB.

The SNOMED CT *Is a* relationship is mapped to the HL7 *componentOf ActRelationship*, just as in Table 1. Other SNOMED CT relationships, such as *Finding Site*, are mapped by the two concepts belonging to the same HL7 *Observation* instance, as described in the example above. A summary of all mappings between HL7 and SNOMED CT is shown in Table 2. The new mappings are shown in bold.

Table 2. Mappings of concepts and relationships between HL7 and SNOMED CT.

Concepts	
HL7 Observation Field	SNOMED CT Subset
code	Observable Entities
methodCode	Procedures
targetSiteCode	Body Structures
interpretationCode	Clinical Findings
Relationships	
HL7	SNOMED CT
component	Subtype
componentOf	Is a
code fields in a single Act	Finding Site, etc

2. Results

2.1. Messaging in the Health Service Bus

HL7 classes have been modeled and implemented in software and can be generated based on a user's selection of SNOMED CT *Observable Entity* concepts and entering of associated data. The clinician (as the user) can enter in patient details and clinical notes for an entire episode of care into a front-end application. These notes are stored in the HL7 *PatientCare Provision* text field, and are optional. The clinician can then add any number of *ObservationEvents* to the care episode. For each *ObservationEvent*, the clinician chooses a corresponding SNOMED CT concept from a list and can enter a value and text. The value can be of any type, and will be interpreted based on the SNOMED CT concept chosen.

A dynamic message model will then be created, which can be traversed recursively and an XML message output. These messages are then communicated between applications along the HSB using JavaSpaces. Synthetic message generation allows the scalability of the HSB architecture to be tested.

2.2. Terminology Server in the Health Service Bus

The SNOMED CT 'Vital Signs' subset has been converted into XML (SNOMED CT is distributed as a tab delimited text file) for the purposes of using it in the Health Service Bus. This is stored in an eXist [7] XML database and allows for great querying power against the terminology. A Jini service acts as a front end to the terminology server, allowing other applications to communicate with it using the HSB.

2.3. Summary of Results

Figure 1 shows the HSB architecture. Large steps have been made towards this goal, namely the communication (the "bus" in the middle of Figure 1) has been implemented in JavaSpaces and allows any PC (regardless of operating system) to communicate with any other using HL7 messages, and the Terminology Server has been implemented as a SNOMED CT XML database with a Jini Service front end. A translation service

between HL7 V2.3.1 (the most popular version of HL7 2 currently in use) and HL7 V3 is currently being worked on, to allow legacy systems to communicate on the HSB.

3. Conclusion

The information modeling strategies outlined in this paper, combined with Jini technologies, and in particular JavaSpaces, form the basis for communication in the Health Service Bus, and pave the way for semantic interoperability in healthcare.

Synthetic message generation allows the scalability of the HSB architecture to be tested and the plan is to simulate the messaging environment of a large hospital or regional health service.

An XML terminology server and a user-interface for entering patients' vital signs, which generates and sends HL7 messages, have been implemented.

The next step is to create translation services so that legacy applications and computers connected to the HSB can also communicate with each other. This is an important step towards seamless communication, and will be a further step towards the goal of total semantic interoperability in healthcare.

Acknowledgments

This material includes SNOMED Clinical Terms® (SNOMED CT®) which is used by permission of the International Health Terminology Standards Development Organisation (IHTSDO). All rights reserved. SNOMED CT® was originally created by The College of American Pathologists. "SNOMED" and "SNOMED CT" are registered trademarks of the IHTSDO.

The ePOC PDA Project is an Australian Research Council Linkage Grant. The authors acknowledge the support of the ARC, UOW, Pen Computer Systems, TACT and South Eastern Sydney & Illawarra Area Health Service (SESIAHS).

References

- [1] IHTSDO, *History: SNOMED*, <http://www.ihtsdo.org/about-us/history/snomed/>, 2007. Last Accessed: 29/10/2007.
- [2] Health Level 7, *About HL7*, <http://www.hl7.org/>, 2007. Last Accessed: 29/10/2007.
- [3] A. Ryan, P. Eklund and B. Esler, "Toward the Interoperability of HL7 v3 and SNOMED CT: a case study modeling mobile clinical treatment", in *MedInfo 2007*: IOS Press, Amsterdam, 2007.
- [4] D.A. Chappell, *Enterprise Service Bus*: O'Reilly, 2004.
- [5] Jini.org, *The Community Resource for Jini Technology*, <http://jini.org/>, 2008. Last Accessed: 14/02/2008.
- [6] J. Sargent, P. Eklund, A. Ryan, et al, "Mobile Information Access and Diffusion in Ambulatory Care Service Settings", in *Proceedings of HIC 2006*, Sydney, Australia, 2006.
- [7] eXist, *Open Source Native XML Database*, <http://exist.sourceforge.net/>, 2007. Last Accessed: 29/10/2007.

The Adoption of IT Security Standards in a Healthcare Environment

Rui GOMES^{a,c} and Luís Velez LAPÃO^{b,c,1}

^a *Hospital S. Sebastião – Sta. Maria da Feira, Portugal,*

^b *Healthcare Systems Group, Instituto Nacional de Administração, Portugal,*

^c *CINTESIS-Faculdade de Medicina, Universidade do Porto, Portugal*

Abstract. Security is a vital part of daily life to Hospitals that need to ensure that the information is adequately secured. In Portugal, more CIOs are seeking that their hospital IS departments are properly protecting information assets from security threats. It is imperative to take necessary measures to ensure risk management and business continuity. Security management certification provides just such a guarantee, increasing patient and partner confidence. This paper introduces one best practice for implementing four security controls in a hospital datacenter infrastructure (ISO27002), and describes the security assessment for implementing such controls.

Keywords: ISO 27002, Security standards, CIO, Healthcare Information Management.

Introduction

Healthcare services aims at serving people well. In this regard, the need for IS standards is recurrent since every year thousands of people died as the result of clinical errors caused by fatigue or inaccuracy that could have been prevented with proper technology [1]. Most of the problems have to do with lack of coordination between systems due to the use of different standards [2]. Anybody waiting for the standards bodies before implementing IS will be waiting such long time, but information security must stay manageable and able to let preventing threats, reduce vulnerabilities and risks. Hospital S. Sebastião (HSS) is aware of the significance of information security issues and the relevance of standards and frameworks such as *Committee Of Sponsoring Organizations of the Treadway Commission* (COSO) [3] for financial processes control, COBIT [4] for information technology (IT) control, “*Health Insurance Portability and Accountability Act*” (HIPAA) [5] to insurance protection and promoting communications standards and ISO 27002 [6] to manage the information security. Our approach here will focus on healthcare IT security issues. The COSO is a group of standards that includes different financial and auditing institutions’ functions, while COBIT, Control Objectives for Information and related Technology is a good

¹ Corresponding Author: Prof. Luís Velez Lapão, Healthcare Systems Group, Instituto Nacional de Administração, Palácio dos Marqueses do Pombal, 2784-540 Oeiras, Portugal, email: luis.lapao@ina.pt.

framework for assessing, managing and reducing IT risks. We aim at applying the ISO27002 standards to HSS taking advantage of its comprehensiveness in implementation details. One must recognise that each framework has their own weaknesses and strengths; e.g. ISO27002 has a complete level of security, but does not contain product-oriented measures, such as those used on COBIT [4].

1. The IT Security Standards in the Healthcare Environment

The use of standards can be viewed from legal and IT architecture perspectives [7]. From the legal perspective, there are ranges of standards that either recommends general or specific scenarios in healthcare. In the USA, HIPAA is a legal requisite and comprehensive health information protection policy, which promotes the development of electronic healthcare transactions and specifically addresses the issues of privacy and security for health related information [5]. The security element specifically distinguish the innate problems in using electronic forms of records keeping and the changing nature of the technology upon which such records are recorded, used and stored. HIPAA has suffered many delays but it had a clear impact on services feasibility [8]. From IT standards perspective, we refer to ISO27002 (former British Standard Institute (BSI) 7799-1:1999) to assist in the development of security plans. It is a “Code of Practice” purposeful on high-level security management, revised in 2005 to cover current technology and business practices. ISO 27002 is intended as a common basis and practical guideline for developing organizational security standards and effective security management practices based on 11 main sections. As a code of practice it cannot be used for certification, so another standard has been developed ISO 27001 (information security management system requirements) which is certifiable [9]. This standard specifies the requirements for security implementation that is customizable for individual organizations. ISO standards are only a starting point, as they do not contain widespread information on how security measures should be implemented or maintained. Other standards exist for specific proposes of health information, particularly for use in e-health information exchange, like HL7 [10] developed as a standard for clinical information exchange and based predominantly on the HIPAA guiding principles. In addition, the CEN (European Committee for Standardization) is putting significant effort into development of healthcare information systems security in Europe. However, this has resulted in an assorted range of standards being developed for specific instances of technology use. Many standards do not include sufficient security-related provision and given the complex nature of standards, it has resulted in a large number of providers selling security management solutions for interpretation of the standards and also to explore its implementation.

2. The Process of Adoption of IT Security Standards: The role of the CIO

It is now accepted that healthcare is one of the most complex businesses with a large diversity of types of interactions [11, 12]. The possibility of using IS to support the services delivery also opens new opportunities. Smith [13] and others [14] have proposed that only Information Systems (IS) could bridge the information “chasm”. Interoperability of healthcare systems can play a critical role in this process. The Institute Of Medicine reports [14, 15] identified weaknesses in the design and safety of

healthcare IS whereas interoperability rules' utilization can provide additional pressure to help the proper use of technology in that regard [16]. Both *technical* and *semantic* interoperability require a wide organizational agreement on standards. Both represent huge tasks to be accomplished and require people in the organization to deal with it. Specialized groups such as IHE are pushing the debate and developing interoperability profiles to tighten areas of ambiguity en route to stronger interoperability. The HL7's Electronic Health Record (EHR) group has produced many reports and other materials to guide technology managers towards interoperability. But before going into this sophisticated processes there are many other basic areas that need to be properly covered, being security issues one of them. The human and organizational side of the interoperability has been mostly forgotten [17, 18]. For a long time healthcare process engineering was also not taken very seriously [19]. In order to take advantage of an IS it is necessary a leadership to promote the alignment of business with IS. In this complex environment the role of the Chief Information Officer (CIO) is critical to ensure good focus on organizational specificities. It was recognized that best performing HIS departments were related with department heads that matched CIO attributes [12], like openness to suggestions and excellent relationship with other healthcare professionals; leadership skills, which help them to address challenges; meaningful negotiation skills which are used in their relationships with the vendors, openness to bolder projects with new technologies; etc. Healthcare CIOs are a kind of "special people" that push the organization further through an innovative use of technology [18, 20]. They know that pushing for interoperability will allow the organization to be more productive and less inefficient. Interoperability in an organization can also mean data access safety and security.

3. The Hospital S. Sebastião Information Security Case

HSS is integrated in the National Health Service providing tertiary health care services for all citizens of its geographical area. Built in 1999, it covers an area with 367 000 inhabitants. HSS was chosen to become involved in an innovative management framework, supported by the Ministry of Health, to show the evidence of the improving efficiency of the new framework.

3.1 HSS Information System Architecture

Hospital owns today a unified IS platform that aims to serve not only administrative and management purposes but mainly patients needs, helping professionals doing their job correctly. This middle management application provides approximately 320 physicians and 510 nurses with an integrated view of all clinical information related with the patients, from exams to surgery reports. Since 1999, those physicians create and stores medical records through the hospital's datacenter storage bank. The IS architecture is showed below (Figure 1.), where all the exclusively solutions contribute to grow the datacenter databases on consolidate and concentrated ness philosophy.

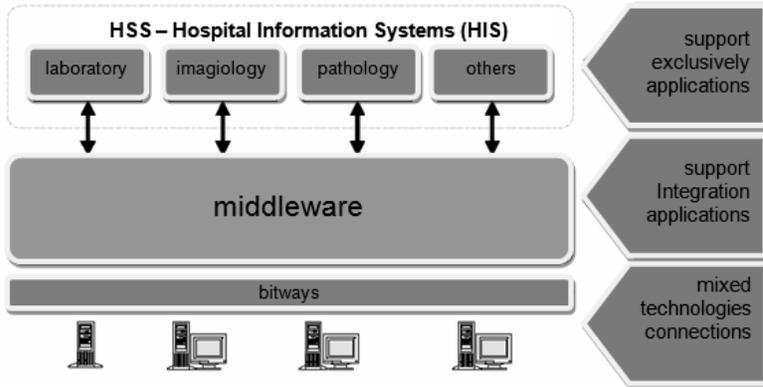


Figure 1. HSS information System Architecture Overview.

The architecture definition was a long working process. The hospital board have recognized that a huge effort was carried out to minimise risks concerning the information management, data privacy and protection. This level of maturity was achieved in 2003, though these good principles are still not enough. These first successes encouraged the CIO, the IT personnel and top managers to be more focused on the improvement of the information security management.

3.2 CIO Role in the HSS IT Security Approach

The CIO created a team to address the IT security at the HSS. After relevant literature and practices review, it was selected the ISO27002 rather than COBIT. COBIT's entirety would make implementation onerous and if one compared it with ISO27002, it is easy to see that it focuses more on efficiency and effectiveness of IT environment rather than information security linked to business issues. It was recognized that ISO27002 represents a good mix of international acceptance level and full comprehensiveness, as well as it is dedicated most exclusively for information security practices built around policy and process management. However, in the future it could be necessary to implement some COBIT measures to accomplish ISO27002 good practices. The applications servers and databases are all concentrated and beneath a controlled physical habitat, and what concerns securing and managing information, the prerequisites surround ISO27002 were recognized as an excellent point of reference to starting managing the information security. Some of controls of this standard have been implemented over a hospital datacenter infrastructure area and the focus has been IT and security policies as a best practice for information security management in the daily basis procedures operation. ISO27002 provides best practice recommendations on information security management for use by those who are responsible for initiating, implementing or maintaining information security management systems. 11 main sections border physical and logical preservation of confidentiality, integrity and availability properties. Making analogy with ISO quality standards and their way of managing and improving hospital made process of ISO27002 implementation as easy as possible. Analogically to quality manager, information security manager observes situation, gives regular assessments, and then recommendations for improvement, afterwards business managers determine to what issues investments should be put in as well as their priority. All 11 ISO27002's control chapters have subset elements. To

provide performance measurement HSS rated the 39 main security categories, based on ISO27002 structures and according to a simple level of risk scale H-M-L (High-Medium/Moderate-Low/Tolerable). The following table 1., concisely shows the risk levels for each control area helping the CIO to rapidly overview the whole picture of information security and to identify priority actions.

Table 1. Risk Levels in the ISO 27002

#	ISO 27002 Section	Risk Level (control objective)		
		H	M	L
1	Security Policy	0	1	0
2	Organizing Information Security	0	1	1
3	Asset Management	2	0	0
4	Human Resources Security	0	1	2
5	Physical and Environmental Security	1	1	0
6	Communications & Operations Management	8	2	0
7	Access Control	5	2	0
8	Information Systems Acquisition, Development and Maintenance	0	4	2
9	Information Security Incident Management	0	2	0
10	Business Continuity Management	0	0	1
11	Compliance	0	1	2

The application of this framework has been quite successful at HSS. For instance, section 3, 5, 6 and 7 were well accomplished in the datacenter infrastructure whereas the security assessment sections were based on the most relevant high-risk level control objective.

3.3 IT Security Project Issues

CEO and CIO have assumed the project and the relevance of a security auditing and its implications: for instance, the obligation to up-grade, both physically and logically, the datacenter, and to change the daily *modus operandi*. The hospital board decided to hire an auditor (named by SINFIC, a BSI certificated partner). The auditor applied a Gap Analysis with five major steps: 1. Project planning, to ensure that expectations, timelines and deliverables are appropriately managed. 2. During the Information-gathering phase many players were interviewed to determine the business environment and current security management and system administration processes through in-depth discussions with key players in the organization. 3. At the Review and Analysis stage Security Policies, Procedures and Practices were addressed to evaluate the existing security policies, procedures and practices, and compare it with the ISO27002 international security standard and industry best practices. 4. The Review and Analysis stage results help to write down a concise, detailed technical and ISO27002 Security Assessment Executive Summary Report. 5. External and Internal vulnerability scanning to discover all devices and applications across the datacenter, and to identify and eliminate the security threats that make datacenter infrastructure attacks possible.

4. Conclusions

From the case presented one should conclude that rules *code of practice* or *standards* are essential to ensure the delivery of benefits to the patient and healthcare providers in information interoperability. This is only part of a bigger effort to implement a comprehensive strategy that allows consistency of information collection and sharing within the healthcare sector. This effort will establish a secure infrastructure between organizations over which to share patient secure information. It is required a comprehensive set of standards that define practical guidelines for the healthcare community, for which ISO27002 is a good benchmarking. Its area of application is a set of diverse and heterogeneous organizations like public hospitals, private, specialists and general practitioners. It means that specific targeted standards should be developed or established for the protection of sensitive information, and not left to individual interested parties to build up. It also means that we are facing a rather new field yet to be proven, implying that the CIO responsible for the implementation of an IS security framework will have to deal with its many variables and barriers. The CIO role and understanding of the organization's environment is key to deliver real interoperability potential to the organization to patients' benefit.

References

- [1] Thompson T. US Former secretary of Health and Human Services Keynote Speech at the 2007 CDHC Expo, Business Wire, Nov. 13, 2006.
- [2] Bell K. "HIT and Pay for Performance". Acting Deputy, US Office of the National Coordinator for Health Information Technology Keynote Speech at the HIT Symposium at MIT, July 17, 2006.
- [3] COSO. "Committee Of Sponsoring Organizations of the Treadway Commission" controls financial processes.
- [4] COBIT. Control Objectives for Information and related Technology", control focuses on IT.
- [5] HIPAA. "Health Insurance Portability and Accountability Act", to insurance protection and promoting communications standards in healthcare, HHS Report 1997.
- [6] ISO/IEC 27002:2005 Information technology -Security techniques - Code of practice for information security management, International Standards Organization, 2005.
- [7] The Role of Standards in Medical Information Security: An Opportunity for Improvement. P. A. H. Williams , School of Computer and Information Science Edith Cowan University Joondalup, Western Australia
- [8] HHS (2007), Health Information Privacy Act (HIPAA). <http://www.hhs.gov/ocr/> (accessed on the 21st October 2007).
- [9] ISO/IEC 27001:2005 Information technology Security techniques - Information security management systems - Requirements, International Standards Organization, ISO/IEC 2005.
- [10] Health Level 7 - ANSI - application layer 7 in the OSI model accredited standards for electronically defining clinical and administrative data in the healthcare industry: www.hl7.org, accessed on the 3rd November 2007.
- [11] Plsek P and Wilson T. Complexity Sciences: Complexity, leadership, and management in healthcare organisations. *BMJ* 2001; 323: 746-9.
- [12] Lapão L.V. Survey on the Status of the Hospital Information Systems in Portugal, *Methods of Information in Medicine*, 2007 46 4: 493-499.
- [13] Smith R. The future of health care systems. *BMJ*, 314:1495 (24 May) 1997.
- [14] IOM Report, Crossing the Quality Chasm: A New Health System for the 21st Century, Institute of Medicine, 2001.
- [15] IOM Report, To Err is Human. Institute of Medicine, 1999.
- [16] Lenz, R and Kuhn, KA. Integration of Heterogeneous and Autonomous Systems in Hospitals, *Data Management & Storage Technology* 2002.
- [17] Lorenzi N and Riley R. *Organizational Aspects of Health Informatics*, Springer-Verlag, 1995.
- [18] Ash JS, Stavri PZ, Kuperman GJ. A Consensus Statement on Considerations for a Successful CPOE Implementation. *J Am Med Inform Assoc.* 2003 May-Jun;10(3):229-34.
- [19] Mango PD, Shapiro LA. Hospitals get serious about operations, *The McKinsey Quarterly* 2001 No.2.
- [20] Broadbent M, Kitzis E. *The New CIO Leader: Setting the Agenda and Delivering Results*. Harvard Business School Press (December 2004).

IHE based Interoperability – Benefits and Challenges

Florian WOZAK^a, Elske AMMENWERTH^b, Alexander HÖRST^b, Peter SÖGNER^a,
Richard MAIR^c, Thomas SCHABETSBERGER^a

^a *University of Health Sciences Medical Informatics and Technology, Research
Division for eHealth and Telemedicine, Hall, Austria*

^b *University of Health Sciences Medical Informatics and Technology, Institute for
Health Information Systems, Hall, Austria*

^c *ITH icoserve information technology for healthcare, Innsbruck, Austria*

Abstract. Introduction: Optimized workflows and communication between institutions involved in a patient's treatment process can lead to improved quality and efficiency in the healthcare sector. Electronic Health Records (EHRs) provide a patient-centered access to clinical data across institutional boundaries supporting the above mentioned aspects. Interoperability is regarded as vital success factor. However a clear definition of interoperability does not exist. The aim of this work is to define and to assess interoperability criteria as required for EHRs. Methods: The definition and assessment of interoperability criteria is supported by the analysis of existing literature and personal experience as well as by discussions with several domain experts. Results: Criteria for interoperability addresses the following aspects: Interfaces, Semantics, Legal and organizational aspects and Security. The Integrating the Healthcare Enterprises initiative (IHE) profiles make a major contribution to these aspects, but they also arise new problems. Flexibility for adoption to different organizational/regional or other specific conditions is missing. Regional or national initiatives should get a possibility to realize their specific needs within the boundaries of IHE profiles. Security so far is an optional element which is one of IHE greatest omissions. An integrated security approach seems to be preferable. Discussion: Irrespective of the so far practical significance of the IHE profiles it appears to be of great importance, that the profiles are constantly checked against practical experiences and are continuously adapted

Keywords. Electronic health record, Interoperability, Health information systems.

Introduction

Optimized workflows and communication between institutions involved in a patient's treatment process can lead to improved quality and efficiency in the healthcare sector. Availability of relevant information can furthermore lead to significant cost reduction and increased patient safety [1, 2, 3]. Electronic Health Records (EHRs) provide a patient-centered access to clinical data across institutional boundaries supporting the above mentioned aspects [4].

Those benefits can however only be exploited if involved systems are able to share required data amongst each other. As a variety of those systems is provided by different manufacturers a vendor-independent interoperability is seen as vital aspect. Integrating the Healthcare Enterprise (IHE) is a non-profit initiative founded in 1998 with working groups in North America and Europe with the aim to stimulate integration of health information systems [5, 6].

Problem Context:

Interoperability is regarded as vital success factor for shared electronic healthcare applications; however a clear definition of interoperability does not exist so far. It is unknown if currently available IHE Integration Profiles sufficiently support interoperability as required for EHRs. It is furthermore unknown if and to which degree extensions have to be applied to existing IHE Integration Profiles to support the desired degree of interoperability.

Objectives

Based on the previously described deficiencies the following goals have been identified:

1. Definition of interoperability as required for EHRs and extraction of assessment criteria.
2. Assessment to which extent the desired degree of interoperability can be achieved by existing IHE profiles.

1. Methods

In an initial step a literature analysis [7] has been carried out to find interoperability requirements for EHR systems. Findings of the literature analysis form the basis for further considerations and are therefore outlined as the first part of the results of this work.

Based on the expertise in the field of Shared Electronic Health Records gained in the health@net project [8] criteria have been elaborated to measure interoperability for EHR systems. Personal experience acquired in the process of integrating IHE compatibility in an EHR system including successful interoperability testing at the 2007 connect-a-thon in Berlin, Germany has been incorporated in the measurement criteria. The interoperability assessment described as objectives is supported by the analysis of existing literature and personal experience. In order to obtain reproducible and valid results those findings have been discussed and aligned with domain experts in the fields of EHR standardization and security as well as with clinical experts which are involved in the requirements definition process.

2. Results

Definition of Interoperability

Interoperability is required to support trans-institutional clinical data exchange in the context of EHRs [9]. Therefore common criteria are necessary which implies a common understanding of the term interoperability. The IEEE for example defines interoperability as the “Ability of a system or a product to work with other systems or products without special effort on the part of the customer” [10].

Criteria for Interoperability Assessment

This definition however addresses the interoperability problem on a very generic level which does not allow the extraction of the desired criteria. A different approach applied in this work is to derive a common set of interoperability criteria from requirements for EHR systems which have been published in literature. In order to fully support trans-institutional exchange of clinical data, basically the same requirements which apply for EHR systems can also be mapped as criteria for their interoperability. In the following enumeration the most relevant interoperability criteria are listed:

- **Interfaces:** Has a common structure of interfaces been agreed on?
- **Semantics:** Are data interpreted identically on systems of different vendors?
- **Legal and organizational aspects:** Are local legal and organizational requirements correctly considered?
- **Security:** Must the security level of a system to be decreased for the sake of interoperability?

The list above has been extracted from requirements for EHRs independently analyzed by [9] and [11]. Interoperable EHR applications must at least support those requirements in order to exploit the above mentioned benefits of trans-institutional clinical data exchange.

Coverage of Interoperability Criteria by IHE Integrating Profiles

In the following section, a discussion of each criterion defined above is given with regard to which extent it can be covered by existing IHE Integration Profiles. The assessment is primarily based on our experience in the field of designing and implementing IHE compliant EHR architectures as well as existing literature.

Interoperability on Interface Level by IHE Integrating Profiles

The core competence of IHE is to analyze common clinical use cases and describe a set of actors and transactions based on existing standards to support those use cases. Systems implemented and tested according to the IHE specifications basically support interoperability on interface level to interconnect EHR applications of different vendors.

However, based on our experience, Integration Profiles leave details open for implementation, which is, from the economic point of view, expected to support the

market. Technically this might lead to slightly different interpretation of the specification, which may prevent successful communication between different systems as experienced in the 2007 connect-a-thon. The concept of IHE is interoperability testing and not certification. This means that systems attested conformity to an Integration Profile can be further developed. This “snapshot” approach allows vendors to further modify and improve their systems even after conformity has been attested, which furthermore allows modifications carried out during the connect-a-thon to be stabilized to a mature system. We however would like create awareness that changes to a successfully tested system might lead to side effects that can, in the worst case, break interoperability.

Interoperability on Semantic Level by IHE Integrating Profiles

Semantic interoperability is a vital aspect to guarantee that data are interpreted identically in trans-institutional EHR applications. IHE has recognized the importance of semantic interoperability and aligned their Integration Profiles with internationally accepted standards for semantic interoperability such as the Clinical Document Architecture (CDA) and common medical dictionaries such as LOINC and SNOMED.

The common data format increases the likelihood that identical meaning of data can be preserved in trans-institutional EHR applications on a technical level. With currently available technology such as CDA, a very fine-grained structure can be applied to clinical data which allows even a language-independent identical interpretation. We however have identified two main challenges which are to be solved in order to profit from those theoretic benefits: (1) As long as there is no strict requirement for producers to apply a fine-grained structure to their data, there will be large amounts of completely unstructured data which are expected to be interpreted anything else then identically. (2) Structuring of clinical data is expected to entirely change the workflow of data capturing. This is an organizational rather than a technical problem as health professionals must be motivated to switch from flexible free text to a rigid framework for data capturing.

Interoperability on Legal and Organizational Level by IHE Integrating Profiles

One of the greatest challenges of the IHE-integration-profiles is the demand to be so specific that it could work, but also to have the corresponding range to fit the legal and organizational requirements of the different nations. The medico legal and organizational scenarios for different areas like Europe, USA or Australia for various actors of healthcare providers differs in many parts, so it's hard for a user group like IHE to develop integration profiles for different actors which meets all requirements worldwide. Therefore the possibility for development of national extensions is a very important precondition.

Interoperability on Security Level by IHE Integrating Profiles

Security is a very complex topic as different institutions with distinct security requirements are involved. The IHE tries to solve security issues by a modular approach which is the definition of additional Integration Profiles such as the Audit Trail and Node Authentication (ATNA) and the Cross Enterprise User Assertion (XUA) which should provide security for existing profiles.

This approach simplifies the implementation process as security can be plugged in at a later time. In our opinion security should however be an integral part of each Profile and not an optional add on. If security becomes an integral part of each application, the concept of application layer security can be implemented which allows authentication and other relevant security features to be handled directly by the communicating applications and not as currently described in the ATNA Profile on network level between communicating nodes.

3. Discussion

So far we have looked at the IHEs potential to solve the problem of interoperability from four different viewpoints. As we figured out the IHE profiles certainly make a major contribution to this problem, but they also arise new problems or do not solve the initial problems completely. Conformity with IHE profiles is not checked continuously by the IHE organization which can lead to interoperability problems itself as systems are constantly enhanced. It seems appropriate to implement a mechanism that deals with this problem. Another problem that is closely associated with IHE profiles is their missing flexibility for adoption to different organizational or other specific conditions. Regional or national initiatives should get a possibility to realize their specific needs within the boundaries of IHE profiles without violation of interoperability. One of IHEs greatest omissions is the lack of security integration in its different profiles. Security so far is an optional element and can only be implemented on network level within the IHE profiles. An integrated security approach seems to be preferable.

Irrespective of the so far practical significance of the IHE profiles it appears to be of great importance, that the profiles are constantly checked against practical experiences and are continuously adapted. Local initiatives like the health@net project [8] are able to provide an important contribution to the enhancement of IHE profiles with their practical experience.

References

- [1] Collen M. General requirements for a Medical Information System (MIS). *Comput Biomed Res.* 1970 Mar;3(5):393–406.
- [2] Ball MJ, Douglas JV. Redefining and improving patient safety. *Methods Inf Med.* 2002;41(4):271–276.
- [3] Maglaveras N, Chouvarda I, Koutkias V, Meletiadis S, Haris K, Balas EA. Information technology can enhance quality in regional health delivery. *Methods Inf Med.* 2002;41(5):393–400.
- [4] Warda F, Noelle G. Telemedizin und eHealth in Deutschland: Materialien und Empfehlungen für eine nationale Telematikplattform. Deutsches Institut für medizinische Dokumentation und Information; 2002.
- [5] Eichelberg M, Aden T, Riesmeier J, Dogac A, Laleci G. A survey and analysis of Electronic Healthcare Record standards. *ACM Computing Surveys.* 2005;37(4):277–315.
- [6] Integrating the Healthcare Enterprise. IHE.net; c2006 [cited 2006 Nov 09]. Available from: <http://www.ihe.net/>.
- [7] Rost F. *Lern- und Arbeitstechniken für das Studium.* 4th ed. Wiesbaden, Germany: VS Verl. für Sozialwissenschaften; 2004.
- [8] Schabetsberger T, Ammenwerth E, Andreatta S, Gratl G, Haux R, Lechleitner G, et al. From a paper-based transmission of discharge summaries to electronic communication in health care regions. *Int J Med Inform.* 2006 Mar;75(3-4):209–215.
- [9] Kalra D. Electronic health record standards. *Methods Inf Med.* 2006;45 Suppl 1:136–144.

- [10] The Role of Standards in Engineering and Technology. Institute of Electrical and Electronics Engineers (IEEE); 2007 [cited 2007 Oct 24]. Available from: http://www.ieee.org/portal/cms_docs/iptorals/iptorals/education/setf/tutorials/baseline/glossary.html.
- [11] Schabetsberger T, Ammenwerth E, Goebel G, Lechleitner G, Penz R, Vogl R, et al. What are Functional Requirements of Future Shared Electronic Health Records? In: Engelbrecht R, Geissbuhler A, Lovis C, Mihalas G. European Notes in Medical Informatics (CD-Rom): Connecting Medical Informatics and Bio-Informatics; 2005.

11. Terminology and Ontology

This page intentionally left blank

Cross-Mapping APACHE IV “Reasons for Intensive Care Admission” Classification to SNOMED CT

F. BAKHSHI-RAIEZ, R. CORNET, N. F. de KEIZER

Department of Medical Informatics,

Academic Medical Center, University of Amsterdam, the Netherlands

Abstract. The APACHE IV classification is used to capture diagnostic information for calculation of mortality risks in intensive care (IC). The lack of structured and formal definitions for concepts in APACHE IV classification as in any classification results in shortcomings when scaling up for re-use. The use of SNOMED CT as a reference terminology can address these shortcomings. However, all of SNOMED CT contains large amounts of information that is irrelevant for IC. By building an interface terminology (Ift) based on SNOMED CT and APACHE IV, it is possible to isolate the IC users from the complexity of SNOMED CT while enabling standardized data registration. Within this study, a mapping is realized from the APACHE IV classification to SNOMED CT. The results of the mapping will be used to identify a relevant SNOMED CT subset for the development of an IC-specific Ift. The vast majority of the diagnostic categories in APACHE IV could be mapped to one or more SNOMED CT concepts (83.8%) and for the remaining concept a partial match was identified (16%). The good mapping results will provide a SNOMED CT subset sufficient for developing an IC-specific Ift. Finally, lessons learned in this study are valuable for other researchers who intend to realize a mapping from a classification to SNOMED CT.

Keywords: SNOMED CT, APACHE IV, mapping, reference terminology, interface terminology, intensive care

Introduction

Healthcare providers documenting clinical encounters are increasingly using classifications to categorize patients with certain diseases or treatments. For instance, calculation of mortality risks in Intensive Care (IC) requires diagnostic information which is captured using the APACHE IV reasons for IC admission classification (APACHE IV classification) [1]. A classification is defined as a systematic arrangement of concepts into classes or groups based on common characteristics [2]. The lack of structure and formal semantic definitions for concepts in most classifications results in shortcomings when scaling up for re-use for multiple purposes. Accordingly, classifications such as APACHE IV generally serve a specific purpose within a certain medical domain and are mostly unsuitable for the broad registration of daily care processes.

Reference Terminologies (RT) are often used to address these scaling and re-use issues. A RT is defined as a set of concepts and relationships which provide a common reference point for comparisons and aggregation of data about the entire healthcare process and enables consistent and computer-readable coding of clinical data, which is a central feature for the use and re-use of information [3;4]. SNOMED CT is regarded as by far the most comprehensive reference terminology for coding clinical data [5]. SNOMED CT is comprehensive on its own, but also maps to other medical terminologies and classification systems already in use. This avoids duplicate data capture, while facilitating enhanced health reporting, billing and statistical analysis.

In the IC, for instance, SNOMED CT can be used to capture diagnostic information which will not only aid the calculation of mortality risks, but will also facilitate sharing and aggregation of data for other purposes. However, in case all of SNOMED CT is provided to the users, its comprehensiveness forms also its disadvantage in that it contains large amounts of information that is irrelevant for most of the specific clinical disciplines and would, therefore, be very cumbersome to use. IC users, for instance, will only be interested in a small fraction of the concepts included in SNOMED CT. Furthermore, the extensiveness of SNOMED CT for all kind of medical domains does not guarantee that it is sufficient for the collection of detailed information for daily patient care in a specific clinical field such as IC.

Therefore, the use of all of SNOMED CT as an interface terminology (IFT) in specific clinical settings is the subject of discussion in many studies [5-7]. An IFT is a terminology used for systematic collection of clinical data that supports direct entry of these data into electronic medical files. IFTs aim to facilitate the display and collection of clinical data to users in a simple way while simultaneously linking the users' own terms to structured data elements used by specific reference terminologies [4].

We aim at developing an IFT based on SNOMED CT and APACHE IV classification for IC. The local IFT to SNOMED CT will make it possible to overcome the scaling and re-use problems related to APACHE IV classification by taking full advantage of SNOMED CT's semantic structure and description logics. On the same time it becomes possible to isolate the IC users from the complexity of SNOMED CT whose terms and language (i.e. English and Spanish) may be inappropriate and insufficient for the IC and may have an arbitrary level of details. The local IFT will initially be used by clinicians to capture diagnostic information which will be used to derive the APACHE IV diagnostic categories for calculation of mortality risks. Later on the IFT will be gradually expanded to serve multiple purposes such as registration of clinical data for medical record systems and research.

This paper describes the mapping from APACHE IV classification to SNOMED CT. Mapping is defined as "linking the content of the two systems through semantic correspondence" [2]. The results of this mapping will be used to identify the relevant subset in SNOMED CT (i.e. the group of concepts, descriptions, qualifiers and/or relationships relevant for the APACHE IV diagnostic categories) as a first step in developing the IC-specific IFT.

1. Materials and Methods

1.1. SNOMED CT

SNOMED CT is the world's largest concept-based terminological system containing 376,046 medical concepts, associated with 1,060,424 description terms for these concepts, and related to each other by a hierarchy consisting of about 1,359,435 relationships (July 2007 release) [8]. Each concept is uniquely identified and can have multiple descriptions. SNOMED CT has a dynamic organization and is compositional (i.e. it supports 'post-coordination'). Relationships are used to define concepts and to specify how they can be refined or qualified.

1.2. APACHE IV Reasons for Admission Classification

The APACHE IV prognostic model is increasingly applied within the field of IC to adjust observed raw mortalities for case mix differences (severity of disturbance in physiological parameters and the primary reason for IC admission) in order to assess the quality of health care [1]. The relating primary reasons for IC admission are collected based on the APACHE IV classification, which contains 445 diagnostic categories belonging to 116 divisions [1]. Each reason for IC admission is first classified as non-operative or post-operative, next by body system or a transplant or trauma-related category, and then by diagnosis. A residual "other" category is used for unlisted diagnoses within each body system, transplant, and trauma category.

1.3. Mapping Procedure

Each APACHE IV diagnostic category was mapped to SNOMED CT concepts. Composite diagnostic categories (i.e. containing more than one diagnosis such as "chest/spinal trauma") are split in atomic diagnoses which are then mapped to more than one SNOMED CT concepts. On the other hand, some composite diagnostic categories contain repeating atomic diagnoses. "Chest injury", for instance, is part of 16 diagnostic categories in trauma group. Each such atomic diagnosis was only once mapped to SNOMED CT.

SNOMED CT was searched by the first researcher (FR) experienced in SNOMED CT and APACHE IV classification to find matches for diagnostic categories from the APACHE IV classification. The diagnostic categories which could not be mapped by FR were also searched for by a second researcher (RC), a SNOMED CT expert. In case of unclear definitions for the diagnostic categories, the researchers consulted an intensivist. The concepts in SNOMED CT were navigated using Clue browser version 2006.2.30 [9].

Mapping consisted of three consecutive activities: 1) Interpreting and analyzing the meaning of diagnostic categories. 2) Matching one APACHE IV diagnostic category to one or more SNOMED CT concept(s). The diagnostic categories were first matched to pre-coordinated concepts. In case no pre-coordinated match was available, a post-coordinated match was searched for. Concepts that did not exist in SNOMED CT were eventually matched to the appropriate superordinates. 3) Assessing each matched concept-category pair on how well they matched by marking each concept-category pair as "complete match", "non-match" or "partial match" (i.e. matches to superordinate concepts).

Table 1: Examples of mapping types

Examples	APACHE IV diagnosis category	SNOMED CT concept(s)	Comments
Complete match pre-coordinated	Chest/spinal trauma	262525000 chest injury "+" 90584004 spinal cord injury	A composite diagnostic category is split and each atomic diagnose is mapped to an atomic SNOMED CT concept
Complete match post-coordinated	Surgery for subdural hematoma	410771003 surgical procedure for clinical finding and/or disorder : 363702006 has focus = 95453001 subdural haematoma	Pre-coordinated concept was not found in SNOMED CT
Partial match to a superordinate concept	Coronary Artery Bypass Graft (CABG), redo	232717009 coronary artery bypass graft	Redo is not specified in SNOMED CT. The diagnose is mapped to a superordinate concept
Partial match for category "Other"	Other medical respiratory disorder	50043002 disorder of respiratory system	Other can not be specified in SNOMED CT.
Non match	Other medical neuromuscular disorder		Neuromuscular not specified in SNOMED CT "Other" can not be specified in SNOMED CT

2. Results

The 445 diagnostic categories in APACHE IV classification were mapped to 397 atomic SNOMED CT concepts. Table 1 provides some examples of the different mapping types. Table 2 presents the results of the mapping. SNOMED CT provided complete matches for 83.8% of the diagnostic categories and partial matches for 16.0% of the diagnostic categories. Only for one diagnostic category, "non match" could be identified.

3. Discussion

The aim of this project was to develop a mapping from APACHE IV classification to SNOMED CT. For 83.8% of the diagnostic categories in APACHE IV classification a complete match was identified in SNOMED CT. This percentage is in accordance with other studies which showed that SNOMED CT has a coverage up to 90% for various medical domains [6;10-12].

As pointed by other authors, the use of post-coordination resulted in better matching scores [13], 31% of all complete matches was realized through post-coordination. In this study post-coordination was based on the semantic possibilities of SNOMED CT representation and was not restricted to functional limitations of the Clue interface. For 12 diagnostic categories (e.g. "surgery for complications of previous spinal cord surgery") it was semantically possible to post-coordinate them in

Table 2: Results of the mapping between APACHE IV classification and SNOMED CT.

	Number of categories	Complete match (%)		Partial match (%)		No match (%)
		Pre.	Post.	Pre.	Post.	
Cardiovascular	101	54.5	22.8	19.8	3.0	0.0
Gastro-intestinal	57	52.6	31.6	10.5	5.3	0.0
Genitourinary	35	45.7	14.3	5.7	34.3	0.0
Hematology	18	72.2	11.1	11.1	5.6	0.0
Metabolic	18	88.9	0.0	5.6	5.6	0.0
Musculoskeletal	20	75.0	5.0	5.0	15.0	0.0
Neurologic	54	66.7	20.4	7.4	3.7	1.9
Respiratory	46	67.4	21.7	8.7	2.2	0.0
Transplant	24	45.8	41.7	4.2	8.3	0.0
Trauma	72	47.2	50.0	1.4	1.4	0.0
Total	445	57.8	26.1	9.5	6.5	0.2

SNOMED CT. However, the functionalities of Clue interface did not allow that. Since the results of our mapping aim to identify the relevant subset in SNOMED CT which is independent of the Clue interface, these concepts are marked as complete matches.

The diagnostic categories including the word “Other” (e.g. Other cardiovascular or Other Respiratory- medical) are residual categories in all body systems for conditions that can not be allocated to more specific categories. These categories form the largest part (47.8%) of the “partial match” group. The remaining diagnostic categories for which a partial match is identified are mostly named identities such as assessment scales or medical devices. We believe that these partial matches are also of great importance for the identification of the relevant SNOMED CT subset to develop a sufficient IFT for IC.

Previous studies on mapping a classification to SNOMED CT have shown that the mapping can be influenced by the structure and content characteristics of both systems [2;13]. In the APACHE IV classification some categories are provided with classification rules. The category “Respiratory Arrest” for instance is refined with “without cardiac arrest”, while the category “Cardiac arrest” is refined with “with or without respiratory arrest; for respiratory arrest only see “Respiratory arrest”. While these kind of rules are used in classification systems to make clear what should and what should not belong to a class, they are not included in terminological systems such as SNOMED CT. Since the results of our mapping will be used to identify the relevant subset in SNOMED CT for an IFT we ignored these classification rules included in APACHE IV and included every atomic concept from SNOMED CT in our mapping.

Furthermore, as in most classifications, the semantics of the diagnostic categories in APACHE IV are not formally defined, which makes mapping to other systems difficult. For instance, from the structure of the APACHE IV classification it is not clear what the exact definition of the diagnostic category “Non operative admission heart transplant” is. It can be mapped to the “Planned operative procedure for heart transplantation” which is a preoperative concept or to “cardiac transplant disorder” which is a concept meant to describe possible complications of a previous heart transplantation. Expert consultations revealed that both are possible and therefore, both mappings were included in the current mapping.

4. Conclusion

This study was performed as the first part of a larger study, designed to develop an IfT based on SNOMED CT and APACHE IV classification for IC. The aim of this paper was to describe the mapping from APACHE IV classification to SNOMED CT.

The vast majority of the diagnostic categories in APACHE IV could be mapped to one or more SNOMED CT concepts and for the remaining concept a partial match was identified. We believe that these good mapping results will provide us with a SNOMED CT subset sufficient for the IC-specific IfT. Furthermore, the findings of our study as described in discussion section provide valuable lessons for other researchers who intend to develop an IfT based on SNOMED CT starting with a predefined classification.

In the near future, we intend to build the IC-specific IfT based on this mapping. The IfT will initially be used to capture diagnostic information for calculation of mortality risks. After an evaluation in the IC setting, the IfT will be gradually expanded to a broad and richly-nuanced set of terms (and concepts) that accurately represent phrases and expressions occurring in the IC domain.

References

- [1] J.E.Zimmerman, A.A.Kramer, D.S.McNair, and F.M.Malila, Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients*, *Crit Care Med* 34 (2006) 1297-1310.
- [2] M. Imel, and J.R. Campbell, Mapping from a Clinical Terminology to a Classification. AHIMA's 75th Anniversary National Convention and Exhibit Proceedings .October 2003.
- [3] K.A.Spackman, K.E.Campbell, and R.A.Cote, SNOMED RT: a reference terminology for health care, *Proc.AMIA.Annu.Fall.Symp* (1997) 640-644.
- [4] S.T.Rosenbloom, R.A.Miller, K.B.Johnson, P.L.Elkin, and S.H.Brown, Interface terminologies: facilitating direct entry of clinical data into electronic health record systems, *J.Am.Med.Inform.Assoc.* 13 (2006) 277-288.
- [5] J.R.Campbell, Semantic features of an enterprise interface terminology for SNOMED RT, *Medinfo* 10 (2001) 82-85.
- [6] S.H.Brown, P.L.Elkin, B.A.Bauer, D.Wahner-Roedler, C.S.Husser, Z.Temesgen, S.P.Hardenbrook, E.M.Fielstein, and S.T.Rosenbloom, SNOMED CT: utility for a general medical evaluation template, *AMIA.Annu.Symp.Proc.* (2006) 101-105.
- [7] P.L.Elkin, A.P.Ruggieri, S.H.Brown, J.Buntrock, B.A.Bauer, D.Wahner-Roedler, S.C.Litin, J.Beinborn, K.R.Bailey, and L.Bergstrom, A randomized controlled trial of the accuracy of clinical record retrieval using SNOMED-RT as compared with ICD9-CM, *Proc.AMIA.Symp.* (2001) 159-163.
- [8] The International Health Terminology Standards Development Organisation. Available at <http://www.ihtsdo.org/our-standards/snomed-ct/>. Last accessed: 10th November 2007.
- [9] The Clinical Information Consultancy Ltd. Available at: <http://clininfo.cliniclu.com/home>. Last accessed: November 10th 2007.
- [10] H.Kim, M.R.Harris, G.Savova, and C.G.Chute, Content coverage of SNOMED-CT toward the ICU nursing flowsheets and the acuity indicators, *Stud.Health Technol.Inform.* 122 (2006) 722-726.
- [11] S.K.Nachimuthu and L.M.Lau, Practical issues in using SNOMED CT as a reference terminology, *Medinfo.* 12 (2007) 640-644.
- [12] H.Wasserman and J.Wang, An applied evaluation of SNOMED CT as a clinical vocabulary for the computerized diagnosis and problem list, *AMIA.Annu.Symp.Proc.* (2003) 699-703.
- [13] A.Vikstrom, Y.Skaner, L.E.Strender, and G.H.Nilsson, Mapping the categories of the Swedish primary health care version of ICD-10 to SNOMED CT concepts: rule development and intercoder reliability in a mapping trial, *BMC.Med.Inform.Decis.Mak.* 7 (2007) 9.

Do SNOMED CT Relationships Qualify?

Ronald CORNET^a

*^aDepartment of Medical Informatics, Academic Medical Center, Amsterdam,
The Netherlands*

Abstract. An important feature of SNOMED CT is post-coordination, which is enabled as SNOMED CT specifies refinability of target concepts of relationships. In this paper the use of refinable relationships in SNOMED CT is analysed, and the interplay between refinable and defining relationships. Refinability is used in 10 out of 61 relationships, and the interplay between refinable and defining relationships is low. The results indicate that the usage of refinable relationships can be further enhanced in SNOMED CT.

Keywords. SNOMED CT, nomenclature, relationships

Introduction

Today, SNOMED CT is among the largest clinical healthcare terminological systems. SNOMED CT provides formal definitions for its concepts using *ISA* relationships and attribute relationships¹. These relationships serve three purposes: making semantics explicit; automated classification; and allowing post-coordination.

First, relationships provide a formal way to reflect the semantics of a concept. SNOMED CT does not provide free-text definitions of concepts, but rather it aims at specifying relevant concept properties in a formal manner. For example, in SNOMED CT PNEUMONITIS is defined as a DISORDER OF LUNG, of which the ASSOCIATED MORPHOLOGY is INFLAMMATION, and the FINDING SITE is LUNG STRUCTURE. (In this paper concepts will be written in UPRIGHT SMALL CAPS, relationships in *ITALIC SMALL CAPS*)

Second, based on definitions of concepts, a reasoning engine can perform automated classification, e.g., can infer that PNEUMONITIS is also an INFLAMMATION OF SPECIFIC BODY ORGANS and an INFLAMMATORY DISORDER OF LOWER RESPIRATORY TRACT.

Third, SNOMED CT specifies two facets of relationships, namely the characteristicity and refinability. The characteristicity is “an indication of whether a Relationship specifies a defining characteristic of the source Concept or a possible qualifying characteristic of that Concept”. The characteristicity can be ‘defining’, ‘qualifier’, ‘historical’ (for versioning purposes), or ‘additional’ (to denote transitivity of the PART OF relationship). Relationships presented above for Pneumonitis had the

¹ SNOMED CT[®] Technical Reference Guide – July 2007 International Release

Table 1. Relationships for pneumonitis specified in SNOMED CT.

source	Relationship	target	characteristic	Refinability
pneumonitis	IS A	disorder of lung	Defining	not refinable
pneumonitis	associated morphology	inflammation	Defining	Optional
pneumonitis	finding site	lung structure	Defining	Optional
pneumonitis	severity	severities	Qualifier	Mandatory
pneumonitis	episodicity	episodicities	Qualifier	Mandatory

characteristic type ‘defining’. In addition to these relationships, three ‘qualifier’ relationships are specified: *SEVERITY*, *EPISODICITY* and *CLINICAL COURSE*.

Refinability is “an indication of whether it is possible to refine the target Concept when this relationship is used as a template for clinical data entry”. Refinability can be ‘not refinable’, ‘optional’, or ‘mandatory’. Table 1 presents the relationships for pneumonitis; the ‘defining’ attribute relationships are all optionally refinable (i.e., one may specify a subtype concept), whereas refining the ‘qualifier’ relationships is mandatory, implying that if one qualifies a relationship, a subtype concept of the target concept of relationships must be selected, e.g., not SEVERITIES but a subtype thereof. Non-refinable qualifiers are used to specify that the target concepts can be selected, but not their subtype concepts.

Whereas inheritance of defining relationships is unambiguously defined by the logic-based representation of SNOMED CT, the (procedural) semantics for qualifier relationships leaves room for interpretation. Hence, it needs to be determined how inheritance of qualifier relationships should be dealt with. Furthermore, the interplay between defining characteristics and qualifier characteristics is of importance, as discrepancies between defining relationships and qualifier relationships may occur.

In this paper, the use of relationships in concept definitions in SNOMED CT is investigated and the occurrence and the interplay between defining and qualifier relationships are analyzed.

1. Materials & Methods

1.1. SNOMED CT

The July 2007 release of SNOMED CT (20070731) was used. The core of this release contains three text files. In this release, 376,046 concepts are defined, for which 1,060,424 English-language preferred and synonymous descriptions are provided. The concepts are defined using a total of 1,359,435 relationships.

1.2. Defining and qualifier relationships

Post-coordination is the mechanism of supporting construction of detailed concepts without the need to have all these concepts defined in a terminology (i.e., pre-coordinated). Post-coordination is not only relevant for clinical users or applications that want to provide detailed information about a patient, but also for modelers who maintain SNOMED CT.

The terminological system GALEN, which was based on the GRAIL modeling language [1], was among the first to support both post-coordination and logic-based definitions. To this end, GRAIL used a mechanism called sanctioning. Sanctioning is one of the core principles of GRAIL, implying “that nothing is allowed until it is sanctioned by means of a sanctioning statement”². Sanctioning is performed at two levels, grammatical and sensible. Grammatical sanctions represent general relationships (e.g., CLINICAL FINDING concepts have a *FINDING SITE* which is an ANATOMICAL STRUCTURE). Sensible sanctions represent relationships that can actually be formed (e.g., a FRACTURE has a *FINDING SITE* which is a BONE STRUCTURE). Sanctioning in GRAIL not only plays a role in ontology construction, but it can also play part in the process of construction of user interfaces[2] for data entry.

Unlike GALEN, contemporary systems such as SNOMED CT usually define concepts using description logic (DL)[3], providing frame-like qualifiers[4]. This may result in discrepancy between the defining relationships and qualifier relationships. In [5] an approach was presented to solve this problem of possible inconsistency. This approach involved specifying ‘invariants’ for the ontology that always should hold. One of these invariants is that defining relationships should only be introduced in subclasses of concepts for which the relationship is specified as a qualifier. In the case of PNEUMONITIS, this would mean that the concept disorder of lung specifies *ASSOCIATED MORPHOLOGY* as a qualifier, for which a MORPHOLOGICALLY ALTERED STRUCTURE can be qualified. In this way, post-coordination is maximally supported, and each pre-coordinated concept can also be constructed by means of post-coordination.

1.3. Methods and analyses

The core files of the July 2007 release of SNOMED CT are imported into an MS Access database. Queries are created to perform an analysis of types of relationships, analysis of use of qualification and definition, providing insight in the interplay between them.

2. Results

2.1. Types of relationships

In addition to the *ISA* relationship, SNOMED CT defines 61 relationships for concept definition. The usage of these relationships is shown in Figure 1. The solid line presents (on a log scale) the number of times a relationship is used in SNOMED CT, varying from 86,403 times for *FINDING SITE* to 1 for *TIME ASPECT* and zero for *SUBJECT OF INFORMATION*. The bars in Figure 1 show the proportions of specified relationships that are defining relationships and various types of qualifier relations.

² <http://www.opengalen.org/tutorials/grail/tutorial7.html>

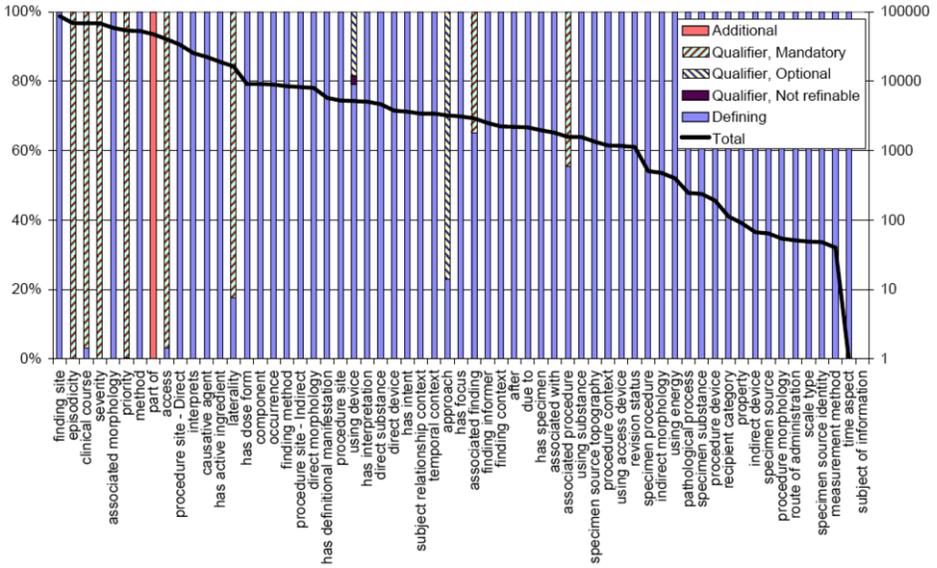


Figure 1. Frequency of use of relationshiptypes (black line), and the proportion of defining and qualifier relationships.

2.2. *Defining relationships*

A total of 923,652 relationships (68%) are defining. The defining relationships that are not refinable are all and only “Is A” relationships (n=499,781). For all defining relationships other than “Is A” refinability is optional.

2.3. *Qualifier relationships*

Fifty relationships are only used for defining relationships. Ten relationshiptypes are used as qualifiers. Of these, the two relationshiptypes *EPISODICITY* and *SEVERITY* are solely used as qualifiers.

As explained in the introduction, qualifier can be non-refinable, optional, or mandatory. There are 131 non-refinable qualifiers, all in the “using device” relationship. For example, the concept BIOPSY OF UVULA AND SOFT PALATE provides a *USING DEVICE* qualifier with values BIOPSY NEEDLE and FINE NEEDLE. These can be selected, but not their subconcepts, e.g., FINE BIOPSY NEEDLE or ACUPUNCTURE NEEDLE.

Optional qualifiers are used for the *USING DEVICE* relationship (n=952) and the *APPROACH* relationship (n=2482). As opposed to the example above, in this case it is allowed to select a subtype concept. For example, for BIOPSY OF LESION OF TONGUE

Table 2. Target concept for mandatory qualifiers and the number of occurrence in SNOMED CT

Relationshiptype	Target concept	#
associated finding	clinical finding	339
associated finding	observable entity	339
associated finding	procedure	339
severity	severities	68095
episodicity	episodicities	68225
access	surgical access	39045
priority	priorities	53336
clinical course	courses	66094
laterality	side	13485
associated procedure	observable entity	351
associated procedure	procedure	351

the value for *USING DEVICE* can be (among others) BIOPSY NEEDLE or FINE NEEDLE or any of their subtype concepts.

Mandatory qualifiers (n=309,999) are used in relationships with generic target concepts. Table 2 shows the target concepts for the relationshiptypes and the number of times these are used. There are 9 different target concepts for all mandatory qualifiers.

2.4. Interplay between defining and qualifier relationships

Out of 61 relationshiptypes specified in SNOMED CT, 1 is never used, 50 are used solely in defining relationships, 2 only for qualifying relationships and 8 are used for both: *CLINICAL COURSE*, *PRIORITY*, *ACCESS*, *LATERALITY*, *USING DEVICE*, *APPROACH*, *ASSOCIATED FINDING* and *ASSOCIATED PROCEDURE*. In these 8 relationshiptypes there is interplay between qualifier and defining relationships.

3. Discussion

First, current problems in the definitions and qualifiers will be discussed, then the advantages of enhancing interaction between definitions and qualifiers, after which further work will be described

3.1. Current problems in the model

The results reveal that in many cases the target concepts of qualifier relationships are generic, such as *SEVERITIES* or *ACCESS METHODS*. Figure 1 shows that the *SEVERITY* relationship is never used as part of definition, which results in the possibility of specifying *SEVERE ASTHMA*, *SEVERITY MILD*. It seems that severity should be a defining property for *SEVERE ASTHMA* or *MILD ASTHMA* and other concepts for which the description explicitly mentions severity.

Another drawback of specifying qualifier at a generic level is that it presents irrelevant qualifiers. For example, *PROCEDURE ON FOOT* will present as *ACCESS* among others the possibility of *MINIMAL ACCESS APPROACH VIA FRONTAL SINUS*.

Analysis of qualifiers shows that they are not necessarily inherited by subordinate concepts. The concept *DISEASE* introduces qualifiers severity, episodicity and clinical course; *CONGENITAL DISEASE* however, does not have an episodicity qualifier. Whereas it makes perfect sense to disallow specifying episodicity of congenital disease, it raises the question of the semantics of qualifiers. For example, it turns out that blocking supersedes inheritance for concepts that have multiple supertype concepts, i.e., if one of the supertype concepts blocks a specific qualifier relationship and another supertype does not block the relationship, the blocking prevails.

3.2. Interplay between defining and qualifier relationships

Only eight out of 60 used relationships are used both for qualifier and defining relationships. This means that the vast majority of relationships are introduced as part of the definition instead of as qualifiers. For example, the concept *TRETINOIN* (a pharmaceutical) does not define a qualifier *HAS DOSE FORM* relationship, so one cannot post-coordinate for example *TRETINOIN GEL*, whereas almost all subtype concepts (e.g.,

TRETINOIN 0.025% GEL) have a defining refinable relationship *HAS DOSE FORM*. The target concept for this relationship is CUTANEOUS AND/OR TRANSDERMAL DOSAGE FORM, rather than CUTANEOUS GEL or CUTANEOUS CREAM, which would more precisely capture the semantics.

The fact that all defining relations (other than *IS A*) are optionally refinable does provide a powerful and consistent mechanism for supporting post-coordination.

3.3. Further work

The results presented are based on analysis of the inferred form in which SNOMED CT is distributed. When the stated form is available, more detailed analyses can be performed as the stated form will reflect the view that the modelers have on SNOMED CT. This should also provide insight in the way in which blocking can be made explicit, which is not clear from the distributed model. More insight in the SNOMED CT modeling process and in the tools used will further help understand how modeling is actually performed.

Examples presented in this paper were generated by using the CliniClue¹. CliniClue is the only system known to the author that supports post-coordination. SNOMED CT would benefit from having more systems, and preferably a reference implementation, that fully supports post-coordination.

4. Conclusion

Providing qualifier relationships to support post-coordination is an important feature of SNOMED CT. In the July 2007 release, qualifiers are used in only 10 out of 60 used relationships. The relationships *EPISODICITY* and *SEVERITY* are only used as qualifiers.

There is limited interplay between qualifier and defining relationships. Qualifier relationships are not used as a means of sanctioning defining relationships, as is done in the GRAIL language. SNOMED CT can benefit from further exploring the use of qualifier relationships with proper target concepts to enhance the power of post-coordination and reduce the possibility of post-coordinating insensible concepts.

References

- [1] Rector, A.L., et al., *The GRAIL Concept Modelling Language for Medical Terminology*. *Artif Intell Med* 9 (1997) 139-171.
- [2] Bechhofer, S. and C. Goble. *Using Description Logics to Drive Query Interfaces*. in *1997 International Workshop on Description Logics (DL97)*. 1997. Gif sur Yvette (Paris), 1997.
- [3] Baader, F., et al., *The Description Logic Handbook: Theory, Implementation, and Applications*, ed. F. Baader, et al. 2003, Cambridge, Cambridge University Press, 2003, p. 555.
- [4] Minsky, M., *A framework for representing knowledge*, in *The Psychology of Computer Vision*, P. Winston, Editor. 1975, McGraw-Hill, New York. p. 211-277.
- [5] Cornet, R. and M.C.A. Klein. *Representing and Using Template-Knowledge for a Medical Ontology in Protégé*. in *Protégé Conference*. SMI, Madrid, 2005.
- [6] Cornet, R. and A. Abu-Hanna, *Auditing Description-Logic-based Medical Terminological Systems by Detecting Equivalent Concept Definitions*. *Int J Med Inform*, in press.

¹ <http://www.cliniclue.com/>

Enhancing Knowledge Representations by Ontological Relations

Kerstin DENECKE

Research Center L3S, University of Hannover, Germany

Abstract. Several medical natural language processing (NLP) systems currently base on ontologies that provide the domain knowledge. But, relationships between concepts defined in ontologies as well as relations predefined in a semantic network are widely unused in this context. The objective of this paper is to analyse potentials of using ontological relations to produce correct semantic structures for a medical document automatically and to ameliorate and enrich these structures. Knowledge representations to unstructured medical narratives are generated by means of the method SeReMeD. This approach is based on semantic transformation rules for mapping syntactic information to semantic roles. Contextual relations expressed in natural language are automatically identified and represented in the generated structures. To achieve additional semantic relationships between concepts, the UMLS Medical Semantic Network and relationships between concepts predefined in the UMLS Metathesaurus are used to support the structuring process of SeReMeD. First results show that these relations can enhance and ameliorate the automatically generated semantic structures.

Keywords. Language Representation, Knowledge-based systems, Terminology, Semantic Network, Natural Language Understanding

Introduction

Ontologies in general supply vocabularies and world knowledge that is necessary for clear communication within a certain domain. Medical ontologies supply words or word groups (concepts) that are used in the domain of medicine. Often, they provide information on relationships between the concepts (e.g. SNOMED CT defines an ‘is-a’ relationship between the concepts ‘*appendix*’ and ‘*large intestine part*’).

To gain information from medical narratives, NLP techniques can be applied. Using these methods, natural language text is transformed into standardised and normalised semantic structures from where key information can be extracted and be made available for further applications. Existing medical NLP systems, like MedLEE [1], HITEx [2] or LUPUS [3], normally use medical ontologies only for mapping natural language to corresponding concepts of the terminology. But relationships provided by an ontology are widely unused in medical NLP systems. Baclawski et al. describe in [4] an approach to exploit the UMLS for generating and indexing knowledge representations for medical documents. Their representation rules are based on syntactic information and semantic knowledge from the ontology, respectively.

Nevertheless, a semantic representation for a complete document is not produced by their algorithm.

Therefore, this paper analyses the potentials of using relationships predefined in medical terminologies for semantic representation. In particular, it will be analysed whether relations defined in the UMLS and in the UMLS Medical Semantic Network (UMLS MSN), respectively, can help to avoid errors in structure generation using SeReMeD [5].

1. SeReMeD – Method for generating semantic representations

SeReMeD generates semantic representations for medical documents automatically [5]. Out of an unstructured medical document, relevant concepts are identified and linked to each other according to the semantic expressed in the document. SeReMeD abstains from a deep syntactic analysis. Instead, semantic relations between concepts are derived from semantic patterns.

SeReMeD's processing pipeline consists of several steps: Within a preprocessing step, a document is dissected into paragraphs and is further structured into sentences and segments (e.g. the head segment [*There is no mass identified*] and the prepositional segment [*on this examination*]). Tokens are detected and their morphologic and linguistic properties are determined. By means of regular expressions, each sentence is checked for special expressions (i.e., dates, quantities, expressions with special meanings like *No evidence of*) and negated structures like *No effusions are identified*. This information is used during the semantic analysis.

The semantic analysis first considers each segment separately: By means of MetaMap [6], each segment string is mapped to UMLS concepts. Out of the possible candidates provided by MetaMap, SeReMeD selects the first proposal out of the highest ranked candidate set. Each UMLS concept of a selected proposal consists of its free-textual description, a unique identifier and the UMLS semantic type, the concept belongs to. Furthermore, each semantic concept belongs to an UMLS main category that can be derived from the semantic type (e.g., the semantic type *Body Location or Region* belongs to the UMLS main category *Anatomy*) and may have a semantic role.

Out of the set of semantic concepts generated in the preceding step, SeReMeD then determines for each segment a main information unit based on the assumption that each sentence deals with one semantic entity that is modified by all other entities. In clinical narratives this main entity describes normally a diagnosis / morphological change or a procedure / treatment. All other information is considered as modifying information.

Finally, the concepts have to be related to each other. SeReMeD links each modifying concept to the corresponding main information of the same segment. Furthermore, the main information of the prepositional segments is linked to the main information of the head segment. It is assumed that the information described in prepositional segments modifies the information of the head segment. The relation type of the introduced relation is given by the meaning of the preposition, which is interpreted according to the context it is used in (e.g., the preposition “*with*” can express the meaning *using* or *accompanied by*). In this way, SeReMeD produces for each sentence of a document a conceptual graph-like semantic representation of its relevant medical contents (see example 1).

2. Problems in generating semantic structures

SeReMeD has been evaluated based on 100 radiology reports (800 sentences) randomly selected from the AMIA corpus [13]. 80% of the sentences were represented correctly by SeReMeD. Besides incomplete or incorrect mapping of sentences to UMLS concepts, missing processing of prepositions (22%), coordinated structures (15%), specific syntactic dependencies (e.g. between adjectives and nouns, 17%), other complex syntactic structures (e.g. coreferences, relative clauses, 32%) and a wrong detection of the main information (38%) led to incorrect representations of sentences.

Example 1

(1) *Mediastinal structures show no evidence of mass.*

(NEG) (Finding) *Mass of body structure*

|- (Body Location or Region) *Mediastinum*

|- (Spatial Concept) *Structure*

(2) *Pleuritic pain on the left.*

(Sign or Symptom) *Pleuritic pain*

|- (localisation) (Functional Concept) *left*

Example 2:

For the phrase *Inconspicuous representation of the pancreas organ* SeReMeD produces:

(Conceptual Entity) *Representation Component*

|- (topology) (Body Part, Organ, or Organ Component) *Entire pancreas*

|- (topology) (Body Part, Organ, or Organ Component) *Body Organ*

The two modifying concepts [*Body Part: Entire pancreas*] and [*Body Part: Body Organ*] are not correctly linked. Instead, one additional relation is introduced, that is the relation between [*Conceptual Entity: Representation Component*] and [*Body Part: Body Organ*].

These results suggest that SeReMeD is able to represent especially phrases with limited complexity correctly (see example 1), e.g. simple noun phrases or noun phrases with a single prepositional phrase attached. For complex syntactic-semantic structures (e.g., coordinated structures, sentences with references) the produced structures are often faulty (see example 2). SeReMeD originates from the assumption that the main information of a sentence is described in the first segment. The main information is not detected correctly if another segment contains this main information. Dependencies beyond main / modifying information - relations remain also unconsidered. Even relations between two modifying information or two main information are not represented (see example 2). For this purpose, additional syntactic or semantic information is necessary. In the following, the latter is considered, i.e. potentials of supporting SeReMeD by means of relations provided by the UMLS Metathesaurus and the UMLS Medical Semantic Network are analysed. The aim is to generate more detailed semantic structures and to identify dependencies hidden for SeReMeD.

3. Enrich semantic structures by predefined semantic relationships

3.1. Ontological relationships within the UMLS

The UMLS (<http://umlsinfo.nlm.nih.gov/>) combines many medical vocabularies and provides a mapping structure between them. The UMLS Metathesaurus provides biomedical vocabularies that in turn contribute thesaural relationships between concepts (e.g., 'is-a' - relationships). The UMLS MSN [7] is a network of general semantic categories or types. It provides 134 semantic types that are grouped in turn into 15 semantic groups (e.g. the concept *atrial fibrillation* belongs to the semantic

types *Finding* and *Pathologic Function* that in turn belong to the group *Disorders*). The MSN defines hierarchical relationships ('*is-a*') and non-hierarchical relations ("*physically related to*", "*spatially related to*", "*temporally related to*").

The relations between semantic concepts ("interconcept relations") and the relationships between semantic types ("intertype relations") offer different possibilities to support SeReMeD. They are introduced in the following sections.

3.2. Interconcept relations

One of the most frequently occurring interconcept relation is the "is-a"- relation, that provides information on inferior and superior concepts. This hierarchical relation can be applied in SeReMeD for checking whether a concept that is selected as main information is subsumed by another concept that is determined as main information. In the same way, dependencies between modifying concepts can be determined. By applying Rule 1 in SeReMeD, hierarchical relations are used while generating semantic representations and redundant concepts are excluded of being represented (example 3).

Rule 1: If the terminology defines an "is-a" - relation between two concepts selected as main concepts or between two modifying concepts of the same segment, the semantic structure will only contain the more specific concept. The generic term will be excluded from representation.

Rule 2: A MSN relation between the semantic types of two concepts of the same segment should be represented by the semantic structure: The concepts are linked to each other according to the relation.

3.3. Intertype relations

Intertype relations are used to detect dependencies between concepts of one segment (example 4). Semantic structures produced by SeReMeD can be enriched and become more detailed by applying Rule 2.

Example 3

Consider the phrase *Inconspicuous representation of the pancreas organ* that is represented in Example 2. Thesaural relations are defined in the Metathesaurus: *[Pancreas] → isA → [digestive organ] → isA → [Body organ]*. Using Rule 1, the concept *[Body Organ]* is excluded from representation because it does not provide any additional information:

(Conceptual Entity) Representation Component

|- (topology) (Body Part, Organ, or Organ Component) Entire pancreas

Example 4:

The sentence *There has been interval increase in the left upper lobe air space disease* is represented by SeReMeD as:

(Idea or Concept) Data Type Interval

|- (localisation) (Body Part, Organ, Organ Component) Structure of left upper lobe of lung

|- (Disease or Syndrome) Disease

(Functional Concept) Increase

|- (localisation) (Body Part, Organ, Organ Component) Structure of left upper lobe of lung

|- (Disease or Syndrome) Disease

The two concepts *[Data Type Interval]* and *[Increase]* are represented as main information and are not correlated within the semantic structure. By applying rule 3 and the semantic network relation *Functional Concept → isa → Idea or Concept* a modified structure can be generated.

(Idea or Concept) Data Type Interval

|- (Functional Concept) Increase

|- (localisation) (Body Part, Organ, Organ Component) Structure of left upper lobe of lung

|- (Disease or Syndrome) Disease

4. Case study

An experimental study shows the potentials of the introduced extensions of SeReMeD by means of ontological relations. The wrong representations produced during the preceding evaluation of SeReMeD (see above) have been analysed. Around 50% of wrongly represented sentences could be corrected using the procedures introduced in this paper. The semantic structures of these sentences contain more than one main information or more than one modifying information for at least one segment. Within the UMLS or UMLS MSN, relations could be identified that could help to generate correct semantic structures for these sentences.

A complete evaluation of the introduced methods is still necessary. But this requires solutions for selecting the “correct” relationship out of the predefined relations. Normally, more than one relation between two semantic types is provided by the UMLS MSN. In this case study, it is assumed that always the correct relation is selected. On that condition, SeReMeD gains additional information out of predefined relations to produce high quality semantic representations.

5. Discussion

In this paper, an approach has been introduced that uses UMLS relations both predefined in the Metathesaurus and in the UMLS MSN to support knowledge representation by SeReMeD. Several medical natural language processing systems are based on the UMLS Metathesaurus, e.g., it provides the domain knowledge for HITEx [3] and for a modified version of MedLEE [8]. But, thesaural relations or relations between semantic types remain normally unconsidered by these systems. One system, MedIE [9] uses relationships provided by the UMLS Metathesaurus to extract relations between terms in a document.

SemRep, a system introduced by *Rindflesch and Aronson* in [10], is comparable to the approach presented in this paper. SemRep uses two types of semantic rules to interpret phrases semantically: Linguistic patterns are matched against corresponding relationships between UMLS MSN semantic types (e.g., the preposition „in“ corresponds to the MSN relation “*part-of*”). The other kind of rules depends on the UMLS semantic types. But in contrast to the approach here, SemRep aims not to determine a semantic structure for a complete sentence. Instead, it provides semantic dependencies between concepts of an input phrase without creating one single representation for a sentence containing all relevant information. Nevertheless, the semantic rules proposed by *Aronson and Rindflesch* could be used within SeReMeD and could help to determine dependencies between prepositional segments.

An ambiguity problem occurs in our approach when several relationships between two semantic types exist. E.g., between the semantic types *Finding* and *Pathologic Function* three different relationships are defined by the MSN (*associated_with*, *evaluation_of*, *manifestation_of*). To resolve this ambiguity, the context of the concepts needs to be considered. This problem can be resolved by means of rules describing the context of concepts or specific semantic types. *Vintar et al.* address this problem in [11] by relation filtering with inverse document frequency or verbal markers that also may be helpful in the context analysed in this paper. Relations introduced by verbs are widely unused in SeReMeD. In [12], *Schutz and Buitelaar* describe a method for extracting relations from text. They propose to use the extracted relations for extending an ontology. In the approach here, verb relationships could additionally enhance the

generated semantic structures. Another difficulty with UMLS MSN relations arises because two concepts are not necessarily linked to each other even when their semantic types are related by an MSN relation. Inserting such a relation into the semantic structure leads to an incorrect representation.

6. Conclusion

In this paper, the potentials of using semantic relations predefined in the UMLS and the UMLS Medical Semantic Network for knowledge representation are analysed. It can be noticed that the resulting semantic structure can be more correctly by considering additional relations. Relationships that are implicitly contained in a sentence can be represented as well. These relations are accessible using world knowledge defined in an ontology. Furthermore, redundant information can be excluded from representation. In this way, they can support the structuring process of SeReMeD especially for handling more complex sentence structures.

Future research will include the filtering of UMLS relations between semantic types to tackle the ambiguity problem. In addition, an evaluation has to be performed to be able to give more meaningful statements regarding the possible improvements of the proposed idea. To avoid superficial relation search and production of incorrect structures, the introduced process should be supervised by the system. This problem is focused in our ongoing work.

References

- [1] Friedman C, Liu H, Shagina L, Johnson S, Hripcsak G: *Evaluating the UMLS as a source of lexical knowledge for medical language processing*. In: AMIA Proceedings 2001; 189-93
- [2] Zeng Q et al.: *Extracting principal diagnosis, co-morbidity and smoking status for asthma research: Evaluation of a natural language processing system*. BMC Med Inform Decis Mak. 2006 Jul 26;6 (1):30
- [3] Schlangen D, Stede M, Bontas EP: *Feeding OWL: Extracting and Representing the content of Pathology Reports*. In: Proceedings of the 4th Workshop on NLP and XML (NLPXML-2004): RDF/RDFS and OWL in Language Technology, Barcelona, Spanien, Juli 2004
- [4] Baclawski K, Cigna J, Kokar MM, Mager P and Indurkha B: *Knowledge Representation and Indexing Using the Unified Medical Language System*. In: Pacific Sym Biocomp, 2000, 490-501
- [5] Denecke K, Bernauer J: *Extracting Specific Medical Data Using Semantic Structures*. In: R. Bellazzi, A. Abu-Hanna, and J. Hunter (Eds.): AIME 2007, LNAI 4594, pp. 257–264, 2007. © Springer-Verlag Berlin Heidelberg 2007
- [6] Aronson A: *Effective Mapping of Biomedical Text to the UMLS Metathesaurus: The MetaMap Program*, AMIA Proceedings, 2001
- [7] McCray AT, Bodenreider O: *A conceptual framework for the biomedical domain*. In: Green R, Bean CA, Myaeng SH, editors. The semantics of relationships: an interdisciplinary perspective. Dordrecht; Boston: Kluwer Academic Publishers; 2002. p. 181-198.
- [8] Campbell DA, Harrington DP, Cimino JJ. *Exploring the semantic relations of radiological terms UMLS resources*. Journal of the AMIA. 2000;7:976
- [9] Zhou X, Han H, Chankai I, Prestrud A, Brooks A: *Approaches to text mining for clinical medical records*. Proc 2006 ACM symposium on Applied computing, Dijon, France, 2006, 235 - 239
- [10] Rindflesch TC, Aronson AR: *Semantic processing in information retrieval*. In: Proc of the 17th Annual Symposium on Computer Applications in Medical Care, 1993, 611-15
- [11] Vintar S, Buitelaar P, Volk M: *Semantic Relations in Concept-Based Cross-Language Medical Information Retrieval*. In: Proc ECML/PKDD Workshop on Adaptive Text Extraction and Mining, Croatia, 2003
- [12] Schutz A, Buitelaar P: *RelExt: A Tool for Relation Extraction in Ontology Extension* In: Proc. of the 4th International Semantic Web Conference, Galway, Ireland, Nov. 2005
- [13] Mamlin BW, Heinze DT, McDonald CJ: *Automated Extraction and Normalization of Findings from Free-Text Radiology Reports*. AMIA Proceedings 2003; 420-424.

Biosurveillance Evaluation of SNOMED CT's Terminology (BEST Trial): Coverage of Chief Complaints

Peter L. ELKIN, MD¹, Steven H. BROWN, MD², Andrew BALAS, MD³, PhD,
Zelalem TEMESGEN, MD¹, Dietlind WAHNER-ROEDLER, MD¹, David
FROEHLING, MD¹, Mark LIEBOW, MD¹, Brett TRUSKO, PhD¹, S. Trent
ROSENBLOOM, MD², Greg POLAND, MD¹

¹*Mayo Clinic College of Medicine*

²*Vanderbilt University*

³*Old Dominion University*

Abstract. The current United States Health Information Technology Standards Panel's interoperability specification for biosurveillance relies heavily on chief complaint data for tracking rates of cases compatible with a case definition for diseases of interest (e.g. Avian Flu). We looked at SNOMED CT to determine how well this large general medical ontology could represent data held in chief complaints. In this experiment we took 50,000 records (Comprehensive Examinations or Limited Examinations from primary care areas at the Mayo Clinic) from December 2003 through February 2005 (Influenza Season). Of these records, 36,097 had non-null Chief Complaints. We randomly selected 1,035 non-null Chief Complaints and two Board-certified internists (one Infectious Diseases specialist and one general internist) reviewed the mappings of the 1,035 chief complaints. Where the reviewers disagreed, a third internist adjudicated. SNOMED CT had a sensitivity of 98.7% for matching clinical terms found in the chief complaint section of the clinical record. The positive predictive value was 97.4%, the negative predictive value was 89.5%, the specificity was 81.0%, the positive likelihood ratio was 5.181 and the negative likelihood ratio was 0.016. We conclude that SNOMED CT and natural language parsing engines can well represent the clinical content of chief complaint fields. Future research should focus on how well the information contained in the chief complaints can be relied upon to provide the basis of a national strategy for biosurveillance. The authors recommend that efforts be made to examine the entire clinical record to determine the level of improvement in the accuracy of biosurveillance that can be achieved if we were to incorporate the entire clinical record into our biosurveillance strategy.

Keywords. Concept representation, Epidemiological Research, Standards

Acknowledgements. This work has been funded in part by grants from the National Library of Medicine LM06918 and the Centers for Disease Control and Prevention PH000022 and HK00014. The authors would like to thank Luka Bajzer for his technical support of this research.

Introduction

As bioterrorism has become a real and increasing threat to our society, early detection holds the promise of limiting the damage resulting from such an attack. Bioterrorism represents a significant threat to the public health of nations. In warfare it is said that if you kill one adversary, you take one person from the battlefield; but if you make one adversary *ill*, you take *many* people from the battlefield, as it takes additional human resources to protect, relocate, and care for the infected individual, even without counting the effects on others that might be secondarily infected. Given the potential of this strategy to further the goals of our enemies, coupled with the minimal infrastructure needed to deliver biologic weapons, the potential for this type of attack represents a serious threat to a countries national security.

One important part of any national defense strategy is to have early detection for outbreaks of illness consistent with exposure to a biologic weapon. An alternate approach, termed electronic biosurveillance, creates a secondary use for information in computer databases that is collected primarily for clinical care and other health-related purposes. In a common approach referred to as syndromic surveillance, data are assigned ("binned") to syndrome categories (e.g., respiratory or gastrointestinal). The data are then analyzed to provide estimates of disease burden by geographic location over time that can be used for early event detection and health situational awareness during an emergency.

BioSense, the U.S. CDC's system for electronic biosurveillance, was initiated in 2003. Data are received from national healthcare organizations including the Department of Defense (DoD), Veterans' Affairs (VA), and others. The ICD-9 diagnosis codes and CPT procedure codes from these sources are binned into 11 syndromes. Data analyses are displayed in the BioSense Application, which can be viewed through the CDC secure data network by CDC, state and local health departments, and DoD and VA personnel. The BioIntelligence Center (BIC) at CDC consists of personnel who monitor the BioSense application daily for data anomalies, assist state and local health department personnel in data interpretation, and contribute to improving the BioSense application.

In mid-2005, CDC began to develop the capacity to receive real-time data from hospital facilities across the United States. This initiative strengthens BioSense by emphasizing access to real-time clinically rich data. Individual patient records are stripped of obvious identifiers such as names, addresses and telephone numbers at the data source before being sent to CDC. Patient populations include those treated in emergency department, hospital inpatient, and hospital-affiliated outpatient settings. Hospital-level data includes a daily hospital census. Patient-level data of interest includes patient chief complaint (*text or Coded Element with no coding system specified*), physician diagnosis[1] (*ICD9-CM codes*), selected ED-specific data (e.g., *vital signs*) (*Numbers or as an Equation*), microbiology laboratory orders (*coded test names, using local coding system*) and results (*LOINC coded tests, SNOMED for Organisms and unspecified coding scheme for other results*), radiology orders (*local codes only*) and results (*Coding system to be determined*), and pharmacy orders (*Drug names are free text; note that in the data model they site RxNorm as one of the standards but it is not actually implemented at present*). As of March 2006, chief

complaint, diagnosis, and selected demographic data were being received from hospitals distributed across the United States.[2]

The Health Information Technology Standards Panel (HITSP)[3] formed a technical committee on biosurveillance in 2006 whose name was later changed to the Public Health Technical Committee. In 2006, the work program produced an interoperability specification focused on the needs of biosurveillance. The specification followed was aimed at satisfying a use case for biosurveillance provided by the American Health Information Community (AHIC) led by Secretary Leavitt.[4] Our work was further informed by a set of data elements thought important to the use case by AHIC and its subgroup charged to develop recommendations regarding biosurveillance data elements.

Given this advice in September of 2006 a first version of the specification was published. Under the current timetable the specification will see some minor revisions and is scheduled to be recognized by Secretary Leavitt by November 2007. Once recognized by the Secretary, companies who wish to do business in this area with the federal government must conform to the specification.

The Certification Commission for Health Information Technology (CCHIT) creates certification criteria for electronic health records. The Interoperability Specifications developed within HITSP are considered input to the certification process and cycle. These standards based specifications are intended to drive interoperability among and between electronic medical record and other Health IT systems.

The Biosurveillance Interoperability Specification relies heavily on chief complaint data for carrying out biosurveillance.[5] As much of the chief complaint data nationally is not recorded in a structured and codified fashion, we chose to investigate how well the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT) covered the concepts in chief complaints found in Mayo Clinic health records.

Method

SNOMED-CT is a large scale clinical nomenclature, which is description logic, based and is produced by the College of American Pathologists. The terminology has greater than 370,000 concepts and 1,000,000 terms. In our lab we add another 790,000 terms as synonyms to make the terminology more clinically friendly (no new concepts are added).

In this experiment we selected 50,000 history and physical records or limited exams from Adult Medicine or Pediatric outpatient visits. All visits occurred in December through February (our typical influenza season). Records were collected in an most recent to oldest fashion starting in 2005 and extending back to 2003 to complete the cohort of 50,000 records. From these source clinical records 13,903 had null or blank chief complaint sections of the record, leaving 36, 097 records with data in the chief complaint field. From these a random sample of 1035 records was selected for review.

All records were processed using the Multi-threaded Clinical Vocabulary Server (MCVS) developed at the Mayo Clinic. The server uses natural language processing to assign free text elements to a controlled representation. Previous research has shown the mappings to be of high accuracy with a sensitivity (recall) of 99.7%, a positive

predictive value (precision) of 99.8% and a specificity of 97.9%.[6][7] Two independent reviewers (a general internist and an Infectious Diseases sub-specialist) reviewed each item. Where the reviewers disagreed a third individual (general internist) adjudicated.

The method for the reviewer judgments (see Figure 1) is outlined as:

1. Is the Chief Complaint term a reasonable term? (Yes--TP or FN; No--TN or FP)
 - A. Yes. Does SNOMED CT match the term exactly?
 - i. Yes: Code a TP for SNOMED CT's ability to code the term
 - ii. No: Look up the SNOMED CT term in a SNOMED CT Browser (MCVS Browser).
 - a. A Match is identified: Code a TP for the SNOMED CT.
 - b. Does not match: Code a FN for the SNOMED CT term.
 - B. No, it is not a reasonable term: Did the MCVS find something in SNOMED CT which maps to this nonsensical term?
 - i. Yes: Code a FP for SNOMED CT.
 - ii. No: Code a TN for SNOMED CT.

CE	Leg ulcers in both the lower extremities.	- Ulcer of lower extremity (disorder) [95344007] [K] - [has Finding Site] . Both lower extremities (body structure) [4180000] [M]
----	---	---

Figure 1: The review tool demonstrating a complete match using a compositional expression.

Final Model of just CC Section To Predict Case Status of Influenza

Symptom ⁺ (In order of best predictor)	Multivariate Model
1) GOLD STANDARD 2=Cough	7.81 (4.12, 14.8) [<.001]
2) GOLD STANDARD 3=Fever	3.90 (2.22, 6.84) [<.001]
3) GOLD STANDARD 9=Infiltrates	0.12 (0.00, 0.74) [0.018]
4) GOLD STANDARD 8=Sore Throat	2.51 (0.77, 8.16) [0.126]
C-statistic	0.714

OR (95% CI) [p-value]

⁺Contribution to regression is mutually exclusive to that of the “best” symptom (and next best symptom, etc.)

Table 1: Final Optimum Case Definition for surveilling Influenza from Chief Complaints

Results:

ROC Curve Final Model for Just CC Section

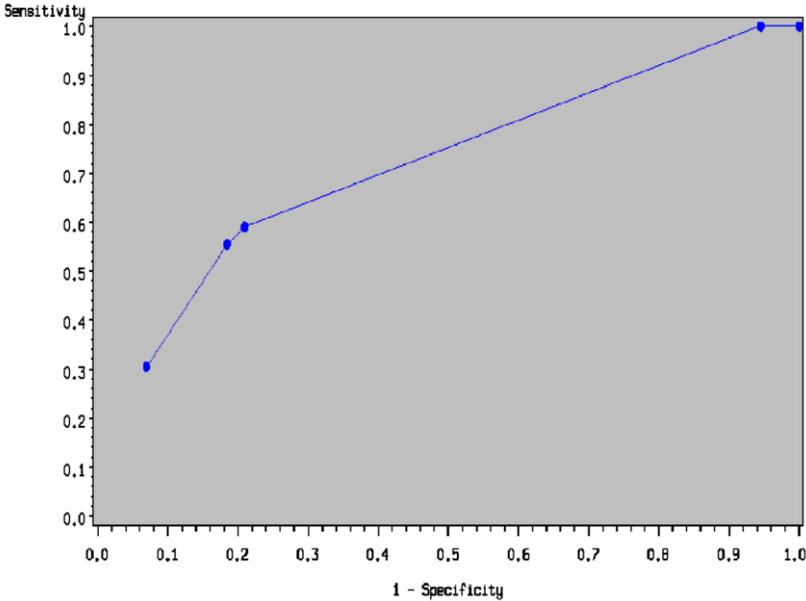


Figure 2: ROC curve representing the various models of Influenza identification from within Chief Complaints.

Overall, SNOMED CT was found to have excellent coverage of the knowledge found in clinical chief complaints. The sensitivity (recall) was 98.7% and the positive predictive value (precision) was 97.4% (see Table 1).

	CE	CID	Partial	Nothing	Total	Total: FP Fixed
tp	452	120	288	0	860	897
tn	16	1	71	6	94	102
fp	11	0	56	2	69	24
fn	3	0	9	0	12	12
Sum	482	121	424	8	1035	1035
Sensitivity	0.993407	1	0.969697	Undefined	0.986239	0.98679868
Specificity	0.592593	1	0.559055	0.75	0.576687	0.80952381
PPV	0.976242	1	0.837209	0	0.925727	0.973941368
NPV	.842105	1	0.8875	1	0.886792	0.894736842
Pos LR	2.438362	Undefined	2.199134	Undefined	2.32981	5.180693069
Neg LR	0.011126	0	0.054204	Undefined	0.023863	0.016307513

Table 2: Results of SNOMED CT mappings of the clinical content coverage of Chief Complaints. Judgments as to the degree of match with the input phrase are generated by MCVS (CID = Concept ID, CE = Compositional Expression). tp = true positive, tn = true negative, fp = false positive, fn = false negative, ppv = positive predictive value and POS LR = positive likelihood ratio).

Some common systematic errors, which were corrected after the first set of reviews, included eliminating proper names from the abstraction and one point of erroneous synonymy (“Pt” mapped to “Pregnancy Detection” instead of “Patient”). The numbers in bold reflect the data after these two sources of error were corrected and the data under total reflect the numbers prior to making these corrections.

Discussion:

Bioterrorism and epidemic infectious diseases such as H5N1 Avian Influenza are a serious and ever present threat to our nation’s public health. In this study, SNOMED CT has demonstrated good coverage for the clinical thoughts / concepts used in chief complaints. It remains unproven as to whether or not the information contained in chief complaints is adequate for the purposes of biosurveillance. A Receiver Operator Characteristics Curve was generated from all the potential models of Influenza surveillance from the chief complaint sections of the record (see Figure 2). A c-Statistic of 0.714 which was obtained from the ROC curve is better than current biosurveillance models for Influenza but not consistent with other common tests used in medical practice such as Weil’s Criteria for Deep Pulmonary Emboli or an exercise stress test which each have c-Statistics around 0.800. This gap raises the question as to whether whole record surveillance is superior to surveillance using chief complaints alone. More research is needed to establish an evidence-based direction in terms of biosurveillance data for our nation. The authors specifically recommend the support of research looking at how data from the entire clinical record can best be utilized to maximize our ability to perform effective biosurveillance.

References:

- [1] Rothschild AS, Lehmann HP, Hripcsak G., Inter-rater agreement in physician-coded problem lists. *AMIA Annu Symp Proc.* 2005;644-8.
- [2] Wetterhall SF., Public health preparedness for bioterrorism. *J Med Assoc Ga.* 2002 Summer;91(2):8-11. Review.
- [3] http://www.ansi.org/standards_activities/standards_boards_panels/hisp/hitsp.asp?menuid=3
- [4] <http://www.hhs.gov/healthit/community/background/>
- [5] Chapman WW, Christensen LM, Wagner MM, Haug PJ, Ivanov O, Dowling JN, Olszewski RT., Classifying free-text triage chief complaints into syndromic categories with natural language processing. *Artif Intell Med.* 2005 Jan;33(1):31-40.
- [6] Elkin PL, Brown SH, Husser CS, Bauer BA, Wahner-Roedler D, Rosenbloom ST, Speroff T., Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists. *Mayo Clin Proc.* 2006 Jun;81(6):741-8.
- [7] Brown SH, Speroff T, Fielstein EM, Bauer BA, Wahner-Roedler DL, Greevy R, Elkin PL., eQuality: electronic quality assessment from narrative clinical reports. *Mayo Clin Proc.* 2006 Nov;81(11):1472-81.

Mining for Adverse Drug Events with Formal Concept Analysis

Alexander ESTACIO-MORENO^{a,1}, Yannick TOUSSAINT^a, Cédric BOUSQUET^{b, c}

^a*LORIA, Campus Scientifique. BP 239, 54506 Vandoeuvre. Nancy Cedex, France*

^b*INSERM UMR_S 872, Eq 20, Faculté de Médecine - Paris 5, France*

^c*University of Saint Etienne, Department of Public Health and Medical Informatics, St-Etienne, France.*

Abstract: The pharmacovigilance databases consist of several case reports involving drugs and adverse events (AEs). Some methods are applied consistently to highlight all signals, i.e. all statistically significant associations between a drug and an AE. These methods are appropriate for verification of more complex relationships involving one or several drug(s) and AE(s) (e.g. syndromes or interactions) but do not address the identification of them. We propose a method for the extraction of these relationships based on Formal Concept Analysis (FCA) associated with disproportionality measures. This method identifies all sets of drugs and AEs which are potential signals, syndromes or interactions. Compared to a previous experience of disproportionality analysis without FCA, the addition of FCA was more efficient for identifying false positives related to concomitant drugs.

Keywords: Adverse Effects, Formal Concept Analysis, Data analysis-extraction tools, Algorithms, Pharmacy.

1. Introduction

Detecting unexpected relationships between drugs and adverse events (AEs) from a pharmacovigilance database is a real challenge. Automated signal detection consists in application of data mining algorithms to help identification of signals, i.e. all statistically significant {drug, AE} pairs, which are numerous and should be reduced to the most plausible [1]. Indeed, these potential signals extracted from a database are only assumptions; they should lead to more complex and expensive pharmacological studies. Therefore, it's important not to ignore any relationships but avoid an excess of studying situations.

Several types of relationships can be identified from the database: (i) {drug, AE} pairs that data mining algorithms have mainly focused on; (ii) Relationships between, two drugs and an AE which are potential drug interactions, (iii) Relationships between several AEs and a drug which are potential syndromes, (iv) Relationships involving several drug(s) and AE(s) which are frequently observed in protocols. Among relationships between several drugs, cases where the drug association strengthens a

¹ Corresponding Author

signal compared to both single drugs may be classified as a potential interaction. Cases where there is no signal between one of the drugs and the AE are more difficult to interpret and may be related to concomitant drugs: the first drug is responsible for the AE but the second drug is only an innocent bystander which was associated to the first drug by chance or due to a frequent therapeutic association.

Automated methods for detecting drug-AE relationships are based on calculation of disproportionality measures. The statistical approach includes the Proportional Reporting Ratio (PRR) [2], χ^2 often coupled with the PRR, and the Reporting Odds Ratio [1]. The Bayesian approach consists of the Multi-item Gamma Poisson Shrinker algorithm [3] and the Bayesian Confidence Propagation Neural Network [4]². Recently, van Puijenbroek *et al.* have used these methods in the detection of drug interactions [5] using a model based on logistic regression. These methods are appropriate for verification of more complex relationships involving several drugs and/or AEs (e.g; syndromes or interactions) but do not address the identification of them. A syndrome detected by case review involving Terbinafine and three AEs (arthralgia, urticaria and fever) was verified using this model [6]. A couple of articles addressed verification of already known drug interactions based on pharmacological knowledge. No systematic research and verification of relationships involving several drugs and/or syndromes was described in the literature with data mining algorithms.

Some applications of Formal Concept Analysis (FCA) have been proposed in the medical field, for example the study of cancer patients flows from DRG databases [7]. We present a new approach that detects all kinds of drug-AE relationships thanks to an exhaustive exploration of a pharmacovigilance database. It helps identifying situations where drugs are prescribed together whereas a previous method [8] generated several false positives due to the innocent bystanders. It is based on FCA [9] combined with disproportionality measures. This article introduces briefly FCA and its use in pharmacovigilance, and then presents the results of an application of this method to an extract of the French pharmacovigilance database. We finally conclude with a discussion of the results.

2. Materials and Methods

2.1. Fundamentals of Formal Concept Analysis (FCA)

The central notion in Formal Concept Analysis [8] is a hierarchical structure called Concept Lattice where objects are classified following the attributes they own. The set of objects (O) and attributes (A) together with their relation to each other ($I \subseteq O \times A$) is called a *formal context* $K = (O; A; I)$. Table in figure 1(a) defines a formal context for adverse drug event exploration, where O is a set of patients and A is the set of attributes: sex (male and female), drugs that patients took (D1, D2, D3, D4, D5) and adverse events (AE1, AE2). Two derivation operators, both denoted by ' link objects and attributes. Let $X \subseteq O$ and $Y \subseteq A$ then $X' = \{a \in A / \forall o \in X, oIa\}$ and $Y' = \{o \in O / \forall a \in Y, oIa\}$. Two compound operators both noted " are defined by the composition of the two ' operators (i.e. " = ' o '). They are *closure operators* over 2^O (" : $O \rightarrow O$) and 2^A (" : $A \rightarrow A$). A pair of '-connected sets is called *formal concept*:

² See [1] for a comparison of these methods

$(X, Y) \in O \times A$ is a formal concept iff $X' = Y$ and $Y' = X$. X and Y are respectively the *extent* and the *intent* of the concept. The set C_K of all concepts of the context K is partially ordered by inclusion \leq_K on 2^O : $(X_1, Y_1) \leq_K (X_2, Y_2) \Leftrightarrow X_1 \subseteq X_2$. $L_K = \langle C_K, \leq_K \rangle$ is a *concept lattice*.

In the context given figure 1(a), $\{P4, P5\}' = \{F\}$ but, following the previous definition, the formal concept built by closure is $(\{P4, P5, P6, P7, P8\}, \{F\})$. The lattice $L_K = \langle C_K, \leq_K \rangle$ has one *top* concept which extent groups all the objects of the context and, by duality, one *bottom* concept which intent groups all the attributes of the context. Objects in the extension of a concept are inherited from the bottom to the top of the lattice. In a dual way, attributes are inherited from top to bottom. Figure 1(b) is the concept lattice of the context (a). The concept $(\{P1, P2, P3, P4\}, \{AE1, D1\})$ is named c_3 in the lattice.

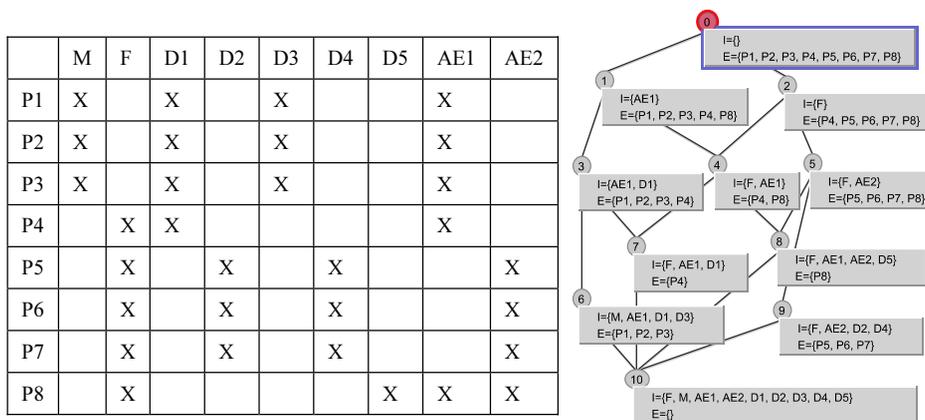


Figure 1: (a) Context describing patients. (b) The corresponding concept lattice

2.2. FCA and Adverse Drug Event Exploration

The intent of some formal concepts shows that some patients took a same set of drugs and had the same set of AEs. In c_3 , four patients $\{P1, P2, P3, P4\}$ took the drug $D1$ and have the adverse event $AE1$. We also observe in the context that $P3$, among others, also took $D3$ but $D1$ is the only drug that the four patient $\{P1, P2, P3, P4\}$ took. Thanks to the closure operators used to build the concepts, the lattice gives all the drug-AE relationships that exist in the data and no more. Among all the concepts, we distinguish potential **signals**, when the intent includes only one drug and one AE (see c_3 and c_7); potential **drug interactions**, when the intent includes two drugs and one AE (see c_6 and c_9); and potential **syndrome**, when the intent includes one drug and several AEs (see c_8).

Compared to extraction of $\{\text{drug, AE}\}$ pairs without FCA, the lattice – with its formal concepts built using the closure operators – reduces the number of drug-AE situations to study. In figure 1(b), patients $\{P1, P2, P3\}$ share the attributes $\{AE1, D1, D3\}$. While disproportionality measures would propose two potential signals to study

($\{AE_1, D_1\}$ and $\{AE_1, D_3\}$), the lattice keeps only one ($\{AE_1, D_1\}$) as $\{AE_1, D_3\}$ is not shared by any group of patients. Of course, thresholds used in disproportionality measures are not meaningful on such a small example.

All formal concepts are not relevant for adverse drug event exploration. Either they do not describe a drug-AE relationship or the described relationship is not statistically significant. To keep statistically significant concepts when the intent includes drugs and AEs, we combine the lattice with criteria used by the British Medicines and Healthcare products Regulatory Agency (MHRA): $\chi^2 > 4$, $PRR > 2$ and the number of patients (the support) is greater or equal to 3 (see [2] for χ^2 and PRR calculation). The a, b, c and d parameters for χ^2 and PRR calculation are calculated from the extent of concepts in the lattice. We tested this method to a subset of 3249 cases from the French database on AEs. Drugs were encoded as active ingredients and AEs were encoded following the WHO-ART terminology. In that way, our database contains 527 drugs and 639 AEs. First, the lattice has been built using the Galicia software and then, for any concept whose support is higher than 3, we calculated both PRR and χ^2 values.

3. Results

The whole lattice has 13,178 concepts. For 842 of them, the intent includes at least one drug and one AE. Relationships are investigated using MHRA criteria in a horizontal strip of the lattice (see black circles in figure 2b). This remaining strip is delimited in the upper part according to the meets, which are the first concepts of the lattice starting from the top including both drugs and AEs in their intent. Delimitation in the lower part is according to the support of concepts because we keep concepts including at least three patients in their extent.

Concepts	Count
Total	13,178
$\{D_1, \dots, D_n; AE_1 \dots AE_m\}$	842
$\{D_1, \dots, D_n; AE_1 \dots AE_m\}$ with $PRR > 2, \chi^2 > 4, \text{Support} \geq 3$	593

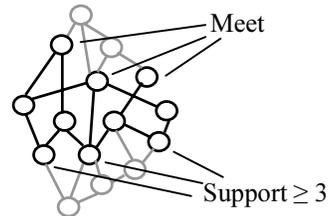


Figure 2: Concepts of the lattice before and after filtering with the MHRA criteria. (a) Results of our application on 3249 pharmacovigilance cases. (b) A graphical representation of conserved concepts (black circles) in a fictitious lattice.

Table 1. Drug-AEs relationships detected with FCA after filtering with the MHRA criteria

Relationship	Count	Type
$\{D_1; AE_1\}$	360	Potential signal
$\{D_1, D_2; AE_1\}$	110	Potential drug interaction
$\{D_1; AE_1 \dots AE_m\}$	56	Potential drug syndrome
$\{D_1, \dots, D_n; AE_1\}$	42	Potential complex interaction
$\{D_1, \dots, D_n; AE_1 \dots AE_m\}$	25	Potential complex syndrome
Total	593	$PRR > 2, \chi^2 > 4, \text{support} \geq 3$

Table 1 shows five different kinds of drug-AEs detected relationship. 181 among the 360 concepts expressing a potential signal are not restricted to a given population (their intent do not include any patient properties, neither age nor sex). We evaluated a sample of 95 of them: 61 signals where already known, for example {Abciximab, Thrombopenia}; 34 were not described in the literature, for example {Lidocaine, Tachycardia} and need further investigations. Among the 56 potential syndromes, 53 consisted of two AEs and three consisted of three AEs.

Table 2. Classification of potential drug interactions

Drug association	Count	Example
Therapeutic associations	57	Fluorouracil + oxaliplatin
Successive prescriptions	8	Heparin + enoxaparin
Medicinal product	22	Amoxicillin + clavulanic acid
Concomitant drugs	23	Heparin + sucralfate
Total	110	

The lattice helps to identify 110 pairs of drugs that might be classified as drug interactions. Among 57 pairs related to drugs commonly prescribed in therapeutic associations, 34 were dedicated to anti-HIV prescriptions. In 8 cases the associated drugs were likely to be prescribed successively rather than at the same time for example heparin was prescribed first then replaced by enoxaparin.

4. Discussion and conclusion

We evaluated in this article the properties of the concept lattice for generating potential relationships between one or several drugs and AEs. Sample size was sufficient for generating several signals. Syndromes mainly consisted of two AEs. No database on syndromes is available to verify their plausibility but we were able to detect known patterns such as hypersensitivity to abacavir. FCA also highlighted more complex relationships that are more difficult to interpret because they are associations of more than two drugs and AEs. This could help the generation of signals related to a whole protocol instead of single drugs. Due to sample size few drug interactions were detected and the interest of the lattice was more focused on the identification of false positives and concomitant drugs.

In a previous experience [8], disproportionality measures extracted 523 statistically significant {drug, AE} pairs with many false positives due to concomitant prescription. The new approach generates only 360 signals. Indeed many signals generated with the previous approach are now clustered in concept nodes of the lattice that present higher order relations, i.e. several drugs and/or AEs. FCA helps to reduce:

1. The number of signals to review. For example the medicinal product consisting of tiemonium + colchicine + opium generated a single signal for the whole association whereas the previous method generated three different signals related to each single active ingredient.
2. The number of false positives related to several active ingredients in the same medicinal products. In previous method a signal of diarrhoea was generated with opium whereas this active ingredient is added in small proportions to

colchicine in order to reduce the diarrhoea episodes. Compound active ingredients are clustered by the current method and it is easier to interpret diarrhoea as related to colchicine.

3. The number of false positives related to concomitant drugs. For example sucralfate was no longer associated to thrombopenia while all patients taking sucralfate were also prescribed heparin.

We showed how FCA may be combined with disproportionality measures for the identification of three kinds of relationships of great interest in pharmacovigilance: signals, drug interactions and syndromes. Building formal concepts by closure with FCA ensures that resulting relationships (signals, interactions and syndromes) are relevant relationships present in data.

The method presented is completely new in pharmacovigilance and, as we saw in the results, finds its validity in signal detection. We are the first to propose a method performing an exhaustive search of syndromes and interactions. No database is available to verifying plausibility of these complex relationships, but results are consistent in terms of pharmacovigilance. We are working on applying an extension of FCA, known as relational concept Analysis, to better characterize the syndromes and interactions.

Acknowledgments

We acknowledge Dr Agnès Lillo-Le Louët, director of the Pharmacovigilance Centre of Georges Pompidou hospital for providing an extract from French Pharmacovigilance database and Dr Pascal Auriche from the *Agence française de sécurité sanitaire des produits de santé* for extracting the data. Alexander Estacio Moreno was supported by a grant from the French « Agence Nationale de la Recherche ».

References

- [1] E.P. van Puijenbroek, A. Bate, H.G. Leufkens, M. Lindquist, R. Orre and A.C. Egberts. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiol Drug Saf.*, **11** (2002), 3–10.
- [2] S.J. Evans, P.C. Waller and S. Davis. Use of proportional reporting ratios (prrs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf.*, **10** (6, 2001), 483–6.
- [3] A. Szarfman, S.G. Machado and R.T. O'Neill. Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the us fda's spontaneous reports database. *Drug Safety*, **25** (6, 2002), 381–392.
- [4] A. Bate, M. Lindquist, I.R. Edwards, S. Olsson, R. Orre, A. Lansner and R.M. De Freitas. A bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol*, **54** (1998), 315–321.
- [5] E.P. van Puijenbroek, A.C. Egberts, R.H.B. Meyboom and H.G.M. Leufkens. Signaling possible drug-drug interactions in a spontaneous reporting system : delay of withdrawal bleeding during concomitant use of oral contraceptives and itraconazole. *Br J Clin Pharmacol*, **47**(6, 1999), 689–93.
- [6] E.P. van Puijenbroek, A.C. Egberts, R.H. Meyboom and H.G. Leufkens. Association between terbinafine and arthralgia, fever and urticaria: symptoms or syndrome? *Pharmacoepidemiol Drug Saf.*, **10**(2, 2001), 135–42.
- [7] N. Jay, A. Napoli and F. Kohler. Cancer Patient Flows Discovery in DRG Databases, in: Ubiquity: Technologies for Better Health in Aging Societies. Proceedings of MIE 2006, Studies in Health Technology and Informatics, IOS Press (2006), 725–730.
- [8] C. Bousquet, C. Sadakhom, C. Le Beller C, M.C. Jaulent MC and A. Lillo-Le Louët A. A rev. of potential signals generated by an automated method on 3324 pharmacovigilance case reports. *Thérapie*, **61**(1, 2006), 39–47.
- [9] B. Ganter and R. Wille. *Formal Concept Analysis. Mathematical Foundations*. Springer, Berlin, 2001.

Automatic acquisition of synonyms from French UMLS for enhanced search of EHRs

Natalia GRABAR^a, Paul-Christophe VAROUTAS^b, Philippe RIZAND^b,
Alain LIVARTOWSKI^b, Thierry HAMON^c

^a *Centre de Recherche des Cordeliers Université Paris Descartes; Inserm, U872; HEGP AP-HP, Paris, France*

^b *Institut Curie, Département d'Information Médicale, 25 rue d'Ulm, 75005 Paris, France*

^c *LIPN-UMR CNRS 7030, Université Paris-Nord, Avenue J.B. Clément, 93430 Villetaneuse, France*

Abstract. Currently, the use of Natural Language Processing (NLP) approaches in order to improve search and exploration of electronic health records (EHRs) within healthcare information systems is not a common practice. One reason for this is the lack of suitable lexical resources: various types of such resources need to be collected or acquired. In this work, we propose a novel method for the acquisition of synonymous resources. This method is language-independent and relies on existence of structured terminologies. It enables to decipher hidden synonymous relations between simple words and terms on the basis of their syntactic analysis and exploitation of their compositionality. Applied to series of synonym terms from the French subset of the UMLS, the method shows 99% precision. The overlap between thus inferred terms and the existing sparse resources of synonyms is very low.

Keywords. Algorithm, Linguistics, Terminology, Data acquisition, Syntax, Lexicons

Introduction

The *Prométhée* framework [1], conceived and deployed at the *Institut Curie* [2], enables the aggregation, cross-interrogation, visualization and simple statistical analysis of heterogeneous biomedical data, including the hospital's electronic health records (EHRs). The framework has been launched in 2002 and is used on a daily basis by the institute's healthcare professionals [3]. *Prométhée*'s current fulltext search engine is limited: users must employ keywords which, at best, would match words and expressions directly used in health records. Guessing such 'best' keywords is a difficult task as the language variation is rich and unpredictable. Within the field of breast cancer, area of expertise of the *Institut Curie*, the International Classification of Diseases for Oncology [4] is used to query structured elements of the EHRs. Even so, its terms do not allow to cover all the expressions occurring in medical records. In this work, we propose a novel method for the acquisition of resources able to improve the

search in French language within *Prométhée*. This method is language-independent and based on the compositionality principle and on the identification of syntactic invariants.

1. Background

Within the NLP field, we distinguish among morphological and synonymous lexica. We notice that these resources are not equally well described for various languages and/or specialized areas. The morphological description of languages is the most complete and several languages are provided with at least inflectional lexica [5-6]. In the biomedical domain, we can mention the widely used lexicon of the UMLS system [7] for English, and similar resources for other languages [8-9]. However, few lexical resources can be found for the description of synonymous relations. Even if WordNet [10] proposes synonym relations for English, the corresponding resources for other languages are not freely available. Moreover, it has been shown that general lexica, for instance WordNet, are insufficient for specialized knowledge extraction [11]. Indeed, additional specialized information is crucial to improve their coverage. In order to provide a solution, we propose to exploit specialized terminologies, as several of them are created and continuously updated in the biomedical area, and decipher the hidden synonyms they contain.

2. Materials

We use the French subset of the UMLS [7] as the original resource from which elementary synonym relations are inferred. The goal of the UMLS is to merge various terminologies (currently over 100) used in the biomedical area. The whole UMLS contains currently over 1,3 million concepts, 71,883 of them have labels in French. Concepts can group several labels in each language, with same or very close meaning. In addition to the UMLS, two available sets of French synonyms were used for comparison purpose: the set of general synonyms comes from the general-language dictionary *Le petit Robert* [12] (140,141 pairs of single-word synonyms); the specialized set of synonyms extracted from the Masson medical dictionary, formerly online [13] (831 pairs of single-word synonyms).

3. Methods: Acquisition of a synonymous lexicon

Within UMLS concepts, terms can show the compositionality through the substitution of one of their components. For instance:

C0016627: {*grippe, influenza, peste*} in *Grippe aviaire, Influenza aviaire, Peste aviaire*

C0016627: {...} in *Avian influenza, Avian influenza, Fowl plague*

C0038352: {*stomacal, gastrique*} in *Contenu gastrique, Contenu stomacal*

C0038352: {*gastric, stomach*} in *Gastric contents, Stomach contents*

We propose a method for the generalization of this observation which allows to acquire a specialized lexicon of elementary synonym relations. In this paper, we shall refer to the series of synonym terms as *original synonym relations*, and refer to the series of their substituted components as *induced* or *elementary synonym relations*. As in the given examples, this method exploits the compositional structure of terms and

relies on existence of structured terminologies. The notion of compositionality, central within this method, assumes that the meaning of a complex expression is fully determined by (1) its syntactic structure, (2) the meaning of its parts and (3) the composition function [14]. In order to exploit this principle, terms are first analyzed syntactically into head and expansion components, then specific inference rules are applied. Finally, the obtained results are evaluated.

Preprocessing of terminology: the Ogmios platform. The aim of the terminology preprocessing step is to provide syntactic analysis of terms. Such analysis is crucial for our work: the proposed method exploits syntactic dependency relations and is based on syntactic invariants. We use the Ogmios platform [15,16], which is suitable for the processing of large amounts of data and, moreover, can be customized for a specialized linguistic domain. First, the TagEN [17] tool is used for the recognition of named entities, in order to help the forthcoming segmentation into words and sentences. Then, POS-tagging and lemmatization was realized with TreeTagger [18]. The step of syntactic parsing of terms is carried with the rule-based term extractor YATEA [19]. The syntactic dependency relations between term components (head and expansion) are computed according to assigned POS tags and parsing rules defined within YATEA.

Acquisition of a synonymous lexicon. The present method is inspired by [14,11]. In [11], authors proposed to apply the semantic compositionality principle for inferring synonymy relations between complex terms. They then postulated that the composition process preserves synonymy and that the compositionality principle holds for complex terms. Roughly, this means that if the meaning M of two complex terms $A \text{ rel } B$ and $A' \text{ rel } B$ are given by the following formulas : $M(A \text{ rel } B) = f(M(A), M(B), M(\text{rel}))$ and $M(A' \text{ rel } B) = f(M(A'), M(B), M(\text{rel}))$ for a given composition function f , and if A and A' are synonymous ($M(A) = M(A')$), then the synonymy of the complex terms can be inferred: $M(A' \text{ rel } B) = f(M(A'), M(B), M(\text{rel})) = f(M(A), M(B), M(\text{rel})) = M(A \text{ rel } B)$. In the current work, we assume that the inverse function f^{-1} exists and, given synonymous complex terms, can be applied for deducing elementary synonym relations. As previously, our approach takes into account the internal structure of the complex terms. We assume that the syntactic dependency relation between components is preserved through the compositionality principle. Thus, we can infer elementary synonym relations between components of two terms if: (1) analyzed terms are synonymous; (2) these components are located at the same syntactic position (head or expansion) and have the same POS tag; (3) the other components within terms are either synonymous or identical. The fully parsed terms are represented as a terminological network, within which the deduction of the elementary synonym relations is based on the three following rules:

Rule 1: If both terms are synonymous and their expansion components are identical, then an elementary synonym relation is inferred between head components. For instance, we can infer the synonym relation $\{\textit{grippe}, \textit{influenza}\}$ (*influenza*) from the original synonym relation between terms *Grippe aviaire* (*Avian influenza*) and *Influenza aviaire* (*Avian influenza*) where the expansion component *aviaire* (*avian*) is identical.

Rule 2: If both terms are synonymous and their head components are identical, then an elementary synonym relation is inferred between expansion components. For instance, we infer the synonym relation $\{\textit{gastrique}, \textit{stomacal}\}$ (*gastric, stomach*) from the synonym relation between terms *Contenu gastrique* (*Gastric contents*) and *Contenu stomacal* (*Stomach contents*) where the head component *contenu* (*contents*) is identical.

Rule 3: If both terms are synonymous and either their head or expansion components are synonymous, then an elementary synonym relation is inferred. For instance, we infer the synonym relation {*pustuleux, vésiculeux*} (*aphthous, vesicular*) from the synonym relation between terms *Angine pustuleuse* (*Aphthous pharyngitis*) and *Pharyngite vésiculeuse* (*Vesicular pharyngitis*) where the head components {*angine, pharyngite*} (*pharyngitis*) are already known synonyms.

Evaluation. We perform manual validation of the inferred elementary relations between terms: each pair is examined, as well as its source series of synonyms. The accuracy of the inferred pairs is thus computed. Moreover, we perform comparison of the acquired synonyms with the existing similar resources: general synonyms from *Le petit Robert* and medical synonyms from *AtMedica* and UMLS and compute the overlap between them.

4. Results and Discussion

Preprocessing of terminology: the Ogmios platform. 156,404 terms, corresponding to 71,883 UMLS concepts have been fully parsed through the Ogmios platform. The original synonym relations were used to infer elementary relations.

Acquisition of a synonym lexicon. The three rules have been applied to the terminological network formed with 76,240 original synonym terms (54,058 original synonym pairs) and generated 1,196 pairs of elementary relations. The general observation is that only three inferred pairs ({*affection, maladie*} (*affection, disease*), {*maladie, syndrome*} (*disease, syndrome*) and {*cancer, tumeur maligne*} (*cancer, malignant tumor*)) are inferred on 10 to 14 series of original synonyms, while the majority of them are supported by singular series of terms. The acquired synonym pairs can be classified according to their linguistic features:

- Orthographic variants: {*acathisie, akathisie*} (*akathisia*), {*embolie, embole*} (*embolus*)
- Abbreviations: {*ARNt, ARN transfert*} (*tRNA, transfer RNA*), {*biop, biopsie*} (*biopsy*), {*EEG, electro-encéphalogramme*} (*EEG, Electroencephalogram*)
- Named entities: {*bartholin, duverney*}, {*côlon, valsalva*}, {*saint jean, rhumatismale*}
- Ellipse: {*insuffisance artérielle, insuffisance*} (*artery insufficiency, insufficiency*)
- Scientific vs popular words: {*maladie, pathologie*} (*disease, pathology*), {*abcès, empyème*}
- Morphologically related words: {*vermiculaire, vermiforme*} (*vermiform*)
- Most induced synonym pairs link entities for which no common formal features can be observed: {*augmentation volume, hypertrophie*} (*enlargement*), {*grave, sévère*} (*severe*), {*cancer, tumeur maligne*} (*cancer, malignant tumor*) ...

Evaluation. Manual evaluation of the totality of generated pairs has shown that 99.3% (n=1188) are correct, 0.08% (n=1) rejected and 0.6% (n=7) not known. The erroneous pair has been generated from the UMLS concept C0038814 and is due to a variation of preposition semantics. But except this pair, the efficiency of the proposed method is very high. This is certainly due to the use of controlled terminological data. Moreover, the inferring rules strongly exploit the syntactic scheme within the syntactically analyzed terms. Finally, these results confirm that French medical terms gathered within UMLS show compositional structure.

As already mentioned, a very large number of elementary synonymous pairs has been deduced on the basis of only one original pair of synonyms. Such small productivity within UMLS can indicate that the inferred pairs are valid in a small number of contexts, but this observation must be verified using other terminologies or corpora. Additionally, the method inferred several pairs composed of named entities (n=26), which use is reduced to the medical area and, possibly, to certain terms only. For instance:

- {*bartholin, duverney*} inferred from C0004768 *Glandes de Bartholin* and *Glandes de Duverney* (*Bartholin Glands, Duverney's gland*),
- {*saint jean, rhumatismale*} inferred from C0152113 *Chorée rhumatismale* and *Chorée de Saint Jean* (*Rheumatic chorea*)

Comparison between the induced elementary synonymous pairs and existing synonyms shows that overlap is very low. Thus, we found only 36 common pairs with the directly available synonyms within UMLS, such as {*tumeur maligne, cancer*} (*malignant tumor, cancer*) or {*saignement, hémorragie*} (*bleeding, hemorrhage*). Thus, the proposed method is useful for it deciphers “hidden” synonyms which are otherwise not accessible. Similarly, we found 2 common pairs with the *AtMedica* resource and 105 with the *Le petit Robert* resource. In the first case, results point out the complementarity between different resources of synonyms from the medical area. As for *Le petit Robert* overlap, the overlap is bigger; still it covers only small part of both resources. It proves that general language resources contain specialized medical vocabulary, although it is not very rich. The difference between them is not surprising as their purpose, as well as addressees and aimed applications, are different. For instance, their use for terminology structuring and knowledge extraction has shown that such general lexica are insufficient for specialized domains [11] and should be completed with specialized resources. Indeed, specialized domains make use of concepts too specific to occur within a general language lexicon.

5. Conclusion and Perspectives

Within healthcare information systems, exploration of EHR content is a current and challenging field. Although the NLP approaches could be suitable to address this need, there is a huge need in various types of linguistic resources. For instance, semantic resources such as synonymous lexica are missing especially in specialized domains. In this paper, we propose a novel method for filling in this gap and inferring synonymous relations between words and simple terms. This method exploits the compositionality principle and relies on existence of structured terminologies. It applies a set of rules based on syntactic dependency analysis within terms. The proposed method has been applied to the UMLS subset of French terms. It provides high-quality results: the manual evaluation showed that over 99% of the inferred relations are correct. The comparison with the available resources of synonyms, such as those directly available in UMLS and sparse resources like *AtMedica* and *Le petit Robert*, is very low. The observed differences seem to indicate that these resources should be combined within NLP tools.

In the near future, we plan to apply the inferred resource to the Institut Curie's EHR corpora (accessible via *Prométhée*), thus further evaluating it during the detection of new synonymous relations between terms used by the institute's healthcare

professionals. We shall then predict their possible impact for users, through analysis of the *Prométhée* query logs. Finally, the detected (and validated) synonyms will be implemented within the *Prométhée* fulltext search engine and possibly by other NLP applications.

As for the method itself, we would like to mention again the fact that it is language-independent. This enables its application to other languages and/or knowledge domains, as long as (1) the required linguistic processing can be realized and (2) synonym relations between complex terms are available.

References

- [1] Varoutas PC, Rizand P, Livartowski A. Using category theory as a basis for a heterogeneous data source search meta-engine: the *Prométhée* framework, *Lect Notes Comput Sci* **4019** (2006), 381-387.
- [2] www.curie.net [accessed 28 February 2008].
- [3] eHealth impact: study on Economic and Productivity Impact of eHealth, commissioned by the European Commission, Directorate General Society and Media. Case study: Institut Curie, Paris, France: Elios and Prométhée, www.ehealth-impact.org [accessed 28 February 2008].
- [4] Fritz AG, Percy C, Jack A, Sobin LH, Parkin MD. International Classification of Diseases for Oncology (ICD-O), WHO, Geneva, Switzerland, 2000.
- [5] Burnage G. CELEX - A Guide for Users. Centre for Lexical Information, Radboud Uni. Nijmegen, 1990.
- [6] Hathout N, Namer F, Dal G. An experimental constructional database: the MorTAL project. In Boucher, P., ed.: Morphology book. Cascadilla Press, Cambridge (MA), 2001.
- [7] NLM: UMLS Knowledge Sources Manual. National Library of Medicine, Bethesda, Maryland. (2007) www.nlm.nih.gov/research/umls/ [accessed 28 February 2008].
- [8] Schulz S, Romacker M, Franz P, Zaiss A, Klar R, Hahn U. Towards a multilingual morpheme thesaurus for medical free-text retrieval. *Stud Health Technol Inform*, **124** (2006), 395-400
- [9] Zweigenbaum P, Baud R, Burgun A, Namer F, Jarrousse E, Grabar N, Ruch P, Duff L, Thirion B, Darmoni S. Towards a Unified Medical Lexicon for French. *Stud Health Technol Inform* **95** (2003), 415-420.
- [10] Fellbaum C. A semantic network of English: the mother of all WordNets. *Computers and Humanities*, **32** (1998), 209-220
- [11] Hamon T, Nazarenko A.: Detection of synonymy links between terms: experiment and results. In: Recent Advances in Computational Terminology. John Benjamins (2001) 185-208
- [12] Le petit Robert. (1990), Paris
- [13] www.AtMedica.com
- [14] Partee BH. In: Compositionality. (1984)
- [15] Hamon T, Nazarenko A, Poibeau T, Aubin S, Derivière J.: A robust linguistic e platform for efficient and domain specific web content analysis. In: RIAO 2007, Pittsburgh, USA (2007)
- [16] www-lipn.univ-paris13.fr/~hamon/ALVIS/Debian/testing, part of the EU project Alvis www.alvis.info
- [17] Berroyer J.F.: TagEN, un analyseur d'entités nommées : conception, développement et évaluation. Mémoire de D.E.A. d'intelligence artificielle, Université Paris-Nord (2004)
- [18] Schmid H.: Probabilistic Part-of-Speech Tagging Using Decision Trees. In: International Conference on New Methods in Language Processing. (1994) 44-49
- [19] Aubin S, Hamon T.: Improving term extraction with terminological resources. In: Advances in Natural Language Processing FinTAL 2006. Number 4139 in LNAI, Springer (2006) 380-387

Design of an automatic coding algorithm for a multi-axial classification in pathology

André HAPPE^{1,2}, Marc CUGGIA², Bruno TURLIN³, Pierre LE BEUX²

¹ *INTERMEDE, La Basse Revachais 35580 GUIGNEN – FRANCE*

² *Laboratoire d'Informatique Médicale 35000 RENNES – FRANCE*

³ *Département d'Anatomie et Cytologie Pathologiques,*

CHU Pontchaillou 35000 RENNES – FRANCE

Abstract. An algorithm for automatic coding of pathology reports using a multi-axial codification (ADICAP) is described and evaluated. It extracts « significant words » or expressions from a corpus and records the statistical relationships between them and the modalities of the different axes. Different weighting functions are evaluated. With the best settings, in more than two cases out of three the correct modality was found among the top 5 list of candidates, except for the « organ » axis. Several ways of improvement are discussed especially regarding the poor results on the « organ » axis. Perspectives of a two stages assembling algorithm completing this first step are proposed.

Keywords. Automatic coding – Pathology – Algorithm - ADICAP Codification

Introduction

Practice guidelines in anatomopathology have been produced for French pathology departments in 1998 [1], specifying that any examined tissue samples should be coded with the ADICAP codification (Association pour le Développement de l'Informatique en Cytologie et Anatomie Pathologique). This multi-axial codification provides lighter handling specifications than ICD 10, it appears less complex than SNOMED [2] and addresses more precisely the sampling processes and tumour grading than ICD-O [3]. Moreover, its availability in French has facilitated its deployment in the French-speaking pathology community in the early 80's. ADICAP codes are composed of 8 mandatory characters [4]: sampling mode (1 character), type of technique applied (1 character), organ (2 characters, the first character represents the organ system) and lesion (4 characters) and 7 optional characters in order to precise the coding [3]. However, manual coding, used in most of the pathology units, is subject to discrepancies between practitioners, lacks of consistency through different structures and requires heavy workload from the coders [5]. This point was emphasized in recommendations for a national survey program of thyroid cancers asking for standardization of diagnosis codification and arguing that an automated coding process would avoid errors or lack of coding [6]. Moreover, such a tool would be useful in the context of the French Cancer Plan which should provide a national information system, allowing interoperability between tumor banks for research purposes and which will

collect coded ADICAP lesions [7]. This paper describes an algorithm that produces automatic ADICAP codes from plain text pathology reports.

1. Material and Methods

1.1. Description of the corpus

Pathology reports produced in 2003 by the Department of Pathology of University Hospital (Rennes – FRANCE) were randomly extracted and anonymised and their mandatory ADICAP codes were collected. The characteristics of the corpus are: 640 reports, 14 practitioners, $396,7 \pm 194,9$ words/report, $1,3 \pm 0,6$ codes/report.

1.2. Description of Wordmapper

Wordmapper [8] is used for the statistical analysis of the reports. It relies on a contextual approach using co-occurrences of words in the text, as highly co-occurring words describe probably a same context. Then, it searches the associations between words in order to identify the « unities of sense » since if two words are often found close to each other they probably bring sense together to the field of study. Though, the expression « lack of sign of malignity » is a unity of sense in itself that outclasses the isolated words « sign », « malignity » or « lack ». Finally, it cross references the « significant words » found in the corpus (isolated or groups of words) with signaletic variables describing each report such as the practitioner or axis of the ADICAP codification. For each variable a contingency table is calculated with « significant words » in lines and the different modalities of the variable found in the corpus in columns. For each cell of the matrix the partial Khi_2 is calculated which represents the contribution of the cell to the global Khi_2 . More the value of the partial Khi_2 is high more the link between the « significant word » to this particular modality of the variable is elevated.

1.3. Treatment of ADICAP codes for Wordmapper

The figure 1 presents an exemple of an ADICAP coded pathology report. For each report, the following signaletic variables are generated : the practitioner producing the report, the activity sector inside the pathology department, the sampling mode(s) of the tissue, the type(s) of technique applied, the organ system(s), the organ(s), presence/absence of either tumoral or non tumoral pathology, presence/absence of a cytopathology, the lesion group(s) for non tumoral pathology, the histogenetic family (ies) and behavior(s) of tumors, the type(s) of studied material for cytopathology. For example, for « Type of technique », we found 3 ADICAP codes with \boxed{E} and 2 with \boxed{E} , so the resulting modality for this axis is $\boxed{E}\boxed{E}$. Moreover, the first code is related to a non tumoral lesion (so Yes, there is a non tumoral pathology), the second and third concern cytopathology (Yes, there is a cytopathology), the two last codes describe tumoral pathology (Yes, there is tumoral pathology). Finally, a descriptor is inserted at the beginning of the report as follows :

[E0300245;XXX;E;O;HE;RS;RPSG;Yes;Yes;Yes;38;E7;0C0M].

Figure 1. The treatment of the ADICAP codes of a report

Informations available	Signaletic variable	Modality	
The report E0300245 is produced by Dr XXXX in the sector E The mandatory ADICAP codes are : [O][E][S][G][38]01 [O][E][R][P][0C]30 [O][H][S][G][0M]30 [O][H][R][P][E7]T0 [O][H][R][P][E7]T3	practician	XXX	
	sector	E	
	Axis of the ADICAP codification	sampling mode	[O]
		technique	[H][E]
		organ system	[R][S]
		organ	[RP][SG]
		non tumoral pathology	Yes
		tumoral pathology	Yes
		cytopathology	Yes
		lesion group	[38]
		family	[E7]
		material	[0C][0M]

1.4. Production of the contingency tables with Wordmapper

A preliminary word frequency analysis is launched on the whole corpus using Wordmapper. It allows to identify the stop words, to ignore all numbers in the reports and to include only words with more than 2 letters whom frequency in the corpus is at least 2. This treatment provides a gross list of 4113 « significant words » (or expressions) which are de-accentuated and capitalized. A manual grouping phase around pivot words is performed that reduces the initial list to 2427 « significant words ». During this stage all proper words and first names are removed. The contingency tables are then produced applying the cross sorting function of Wormapper on each signaletic variable. The result is a matrix for each variable with words in lines and the modalities in columns. The frequency and the partial Khi2 are calculated for each cell of the matrix.

1.5. Weighting functions of the algorithm

In order to modulate the contribution of a « significant word » (or expression) to the ADICAP coding of a specific modality for a given axis, a combined weighting function of the score is used [9]. The weight of the term *t* in the report *c* is expressed as:

$$W(t,c) = W_{local}(c,t) \times W_{global}(t) \times W_{normalization}(c)$$

The algorithm was evaluated using the 3 weighting functions presented in figure 2. In all cases, the local weighting relied on the term frequency (tf factor) since more a term is used in the report more this term is significant for it. A normalization factor was applied only for W_C in order to decrease the influence of term frequency on long reports. For the global weighting two approaches were compared. For a given axis, each time the term *t* was associated in the corpus with a modality of this axis a global weight was evaluated as follows: For W_A an inverted function of the term frequency in the whole corpus was applied so a greater weight is given to rare words in the reports. For W_B and W_C the weight depends on the intensity of the influence of the term *t* on this specific modality for the whole corpus. For exemple, for the axis « Sampling mode » the term « BIOPSIE » (e.g. Biopsy) is associated 9 times in our corpus with the modality « B » (Biopsy) showing a slight attraction ($P < 0.10$), 15 times with the

modality « P » (Punction) with no significative interaction and 158 times with the modality « O » (Surgical piece from radical removal of the organ) resulting in a significative repulsion ($P < 0.01$) as far as the partial Khi^2 are concerned. As high is the significative relationship between the term and the modality as high is the absolute value of the global weight for the term. The sign of the relationship allows to retain the modality as candidate for inclusion if positive (attraction) and for exclusion if negative (repulsion).

Figure 2. Weighting functions used to evaluate the algorithm

Weighting function $W(t,c)$	Local weighting $W_{\text{local}}(c,t)$	Global weighting $W_{\text{global}}(t)$	Normalization $W_{\text{normalization}}(c)$
$W_A(t,c)$	$tf(c,t)$	$\sigma \times \left(10^{\kappa} + \left(\frac{tf(t)}{\max(tf)} \right) \right)$	1
$W_B(t,c)$	$tf(c,t)$	$\sigma \times \log \left(\frac{M}{tf(t)} \right)$	1
$W_C(t,c)$	$tf(c,t)$	$\sigma \times \left(10^{\kappa'} + \left(\frac{tf(t)}{\max(tf)} \right) \right)$	$\frac{1}{n(c)}$

Where $tf(t)$ is the frequency of the term in the whole corpus, σ represents either an significant attraction (+1), repulsion (-1) or non significant influence for a particular modality for a given axis, κ and κ' are factors representing the intensity of the influence of the term t on the modality for a given axis, M is the total number of words in the corpus, $\max(tf)$ is the frequency of the most observed term in the corpus for this given axis and $n(c)$ is the total number of words of the report.

1.6. Evaluation of the algorithm

A randomized sample of 80 coded reports is used for evaluation purpose. They are proposed for coding to the 3 different settings of the algorithm and the modalities proposed by the algorithm are compared to the manual coding. For each report and for each axis a sorted list of the candidate modalities is created on the basis of their cumulated weights. Two indicators are calculated in order to estimate the efficiency of the algorithm: R1 = number of reports where the top candidate modality in the list is correct/total number of reports, R5 = number of reports where one of the top 5 candidate modalities in the list is correct/total number of reports.

2. Results

The table 1 gives the results for each settings of the algorithm. As the indicators of pathology are boolean, the R5 indicator is not mentioned in the table. The results of the $W_B(t,c)$ weighting function are consistently lower than the two other settings. The $W_C(t,c)$ weighting function performance is generally equal or superior to $W_A(t,c)$ for indicator R1 while inferior for R5. For the two best settings, the best performances were observed for the indicator of cytopatholgy, the sampling mode, the material studied for cytopatholgy and the type of technique always in the same order. Except for the organ axis, in more than two cases out of three, the correct modality was found in

the top 5 list of candidates. However, the results on the organ axis remained lower than for all the others axis, the best result gives scarcely one case out of two of correct coding of the organ among the five first modalities proposed.

Table 1. Results of the 3 weighting functions

Axis of the ADICAP codification	$W_A(t,c)$		$W_B(t,c)$		$W_C(t,c)$	
	R1	R5	R1	R5	R1	R5
Sampling mode	77,50%	95,00%	7,50%	20,00%	77,50%	96,25%
Type of technique	55,00%	95,00%	3,75%	22,50%	55,00%	91,25%
Organ	12,50%	51,25%	5,00%	26,25%	12,50%	40,00%
Non tumoral pathology	33,75%	-	30,00%	-	30,00%	-
Tumoral pathology	23,75%	-	23,75%	-	23,75%	-
Cytopathology	78,75%	-	3,75%	-	78,75%	-
Lesion group for non tumoral path.	6,25%	77,50%	7,50%	38,75%	16,25%	66,25%
Family and behavior of tumoral path.	25,00%	76,25%	21,25%	47,50%	31,25%	75,00%
Material studied for cytopathology	67,50%	92,50%	3,75%	13,75%	77,50%	92,50%

3. Discussion

In our study, sequences of words were preferred to isolated words to improve their predictive power for attribution of codes, although for classification purposes another study describe better results with isolated words than with n-grams [10]. This leads to a huge number of items to consider with potential negative effects on performance and therefore filtering rules had to be defined [11]. However, it provided us an elegant way to address the problem of negatives expressions often used in pathology. This point is of concern for all studies on automatic coding, no matter which referential is used as UMLS [12] or SNOMED-CT [13]. The decision to use the multi-axial structure of the ADICAP codification rather than the whole mandatory codes was guided by two observations. First, it was unachievable to build contingency tables between the words of the corpus and the complete ADICAP codes due to their very large number and even so, would have resulted in huge quantities of blank cells in matrix and tiny frequencies in the others. Second, it allows to take into account more precisely the interactions between « significant words » and the axis studied. Thus, the expression « lack of signs » shows a very significant attraction for the modality « Yes » of the indicator of « non tumoral pathology » while in the meantime it seems neutral concerning any modality of « Sampling mode ». Moreover, this approach allows to apply different weighting functions to the different axis of the codification according to their performance. However, the results recorded for the coding of the « organ » axis remained lower than for any other axis no matter the weighting function applied. A first explanation could result from the inverse relationship existing between the number of modalities observed and the performances of the corresponding axis. From a statistical point of view, it could be argued that coding for an axis with less possible modalities is easier than the contrary. Nevertheless, the axis « lesion group for non tumoral pathology » provided 20% better results on R5 than the axis « organ » whereas the number of modalities were of the same order (259 modalities for the organ versus 198). Another reason could come from the content of the reports themselves since 47% of

them involve 2 or more than 2 « organ » codes when 80% of them concern only one « lesion group for non tumoral pathology » code. We plan to adapt the system to characterize the reports related to complex operative pieces resulting from excision of several organs and to introduce this indicator as a signaletic variable. Two ways of improvements will be explored later on; one is a less drastic threshold for candidate words as, for exemple, the very rare expression « physaliphorous cells » was removed in our study while its co-occurrence with specific pathologies (chordoma) has been documented [14]; the other is to treat only the conclusions instead of the whole report.

4. Conclusion

This algorithm could be applied to any multi-axial classification as SNOMED-CT for exemple, provided that a coded corpus is available. The next step of this study will be to assemble the candidate modalities of the different axis to build complete mandatory codes. Two stages will be added to the system, one relies on rules of coding and excludes inadequate associations between modalities of different axis (e.g.: no cytopathology coding when sampling mode is extemporaneous examination). This will help the system to discard innappropriate combinations. In a second step a bayesian network is learned from a gold standard of ADICAP codes. Its nodes are the different axis of the codification and the probabilistic relationships between them are calculated. This will provide guidelines to the system to select the most probable combinations.

References

- [1] J.P. Saint-André et al.,. Recommandations de Bonnes Pratiques en Anatomie et Cytologie Pathologiques. *Annales de Pathologie* **18(3)** (1998); 227-236.
- [2] J. Klossa et al., A European de facto standard for image folders applied to telepathology and teaching, *International Journal of Medical Informatics* **48** (1998); 207-216.
- [3] E. Martin, Thesaurus et codifications des lésions dans les services et laboratoires de pathologie en France, *Gestions Hospitalières* **302** (2001), 59-61.
- [4] F. Boman, A. Petitjean, Proposition de méthode de réponse des frottis cervico-utérins, *Annales de Pathologie* **20(4)** (2000); 390-395.
- [5] AHIMA e-HIM™ Group on Computer-Assisted Coding, Delving into Computer-assisted Coding (AHIMA Practice Brief), *Journal of AHIMA* **75 (10)** (2004), :48 A-H.
- [6] L. Leenhardt, P. Grosclaude, L. Chérié-Challine, Recommandations pour la mise en place d'un dispositif de surveillance épidémiologique nationale des cancers thyroïdiens. Synthèse du rapport, *Institut de veille sanitaire* (2003) 32 pages.
- [7] Institut National du Cancer, LES TUMOROTHEQUES HOSPITALIERES - Recommandations à l'usage des cliniciens et des chercheurs, *Institut National du Cancer* (2006) 121 pages.
- [8] Grimmersoft, WordMapper® Version 8.0 - Analyse et cartographie l'information textuelle, (2004)
- [9] G. Salton, The SMART Retrieval System, Prentice Hall, Englewood Cliffs, NJ., 1971.
- [10] L.M. de Bruijn, A. Hasman, J.W. Arends, Supporting the classification of pathology reports: comparing two information retrieval methods, *Computer Methods and Programs in Biomedicine* **62 (2)** (2000) 109-113.
- [11] J.C. Moissinac, F. Yvon, S. Ben Hazez, Automatic indexing of classes and conferences, *Proceedings of RIAO 2004*, University of Avignon, France (2004) 885-894.
- [12] K.J. Mitchell et al., Implementation and evaluation of a negation tagger in a pipeline-based system for information extract from pathology reports, *Medinfo* **11 (Pt 1)** (2004) 663-667.
- [13] P.L. Elkin et al., A controlled trial of automated classification of negation from clinical notes, *BMC Medical Informatics and Decision Making* **5 (13)** (2005).
- [14] X. Gui, N.H. Siddiqui, M. Guo, Physaliphorous cells in chordoma, *Archives of Pathology and Laboratory Medicine* **128 (12)** (2004) 1457-1458.

Design principles of DOLCE-based formal representation of ICD10

Gergely HÉJA^{a1}, Péter VARGA^b and György SURJÁN^a

^a National Institute for Strategic Health Research

^b Eötvös Loránd University, Institute of Philosophy

Abstract The authors present the framework of formal representation of ICD10 based on DOLCE. The goal of the work is to represent the meaning of the categories of the classification systems based on a formal top-level ontology. The ICD categories are described in the space of atomic disease concepts, while the diseases themselves are defined on a pathological basis. The anatomical entities are taken from FMA, functions, morphological abnormalities and procedures from SNOMED International, while organisms are taken from the biological taxonomy. The ontology is represented in OWL 1.0 in a modularised way. The transformation of the previous GALEN-based representation to the same conceptual system is also under way.

Keywords: Ontology, ICD10, DOLCE, FMA, SNOMED

Introduction

The formal representation of the ICD10 [1] classification system would support the following tasks:

- Terminology based queries of databases: a number of mortality and morbidity databases exist in many countries. Diseases of interest of a user may be scattered throughout the ICD system and manual selection of the relevant codes is error prone. A terminology based query system allows the user to define his/her needs (e.g. all diseases of lung) and a reasoner could automatically collect the relevant ICD codes.
- The 11th revision of ICD is about to be developed, and introduced in a couple of years. Without a mapping from ICD10 to ICD11, the comparability of disease statistics would be lost. Because the huge size, the manual creation of such mapping table is extremely difficult. The formal representation of ICD11 – using the same reference ontology – would enable automatic mapping.

This paper presents the design principles of an ongoing project: the DOLCE [3]

¹ Address for correspondence: National Institute for Strategic Health Research, 6-8. Arany János utca, Budapest, Hungary, H-1051, heja@eski.hu

based formal representation of the meaning of ICD categories. For the creation of the reference ontology, we used existing knowledge sources, described in Section 1. The ontology is described in OWL 1.0 [4], the model is explained in Section 2. This project is a continuation of a previous one [5], in which we defined two chapters of ICD10 based on GALEN [6]. Because of the problems we experienced, and the need for a widespread formal top-level ontology, it has been decided to redesign the ontology on the basis of DOLCE. The transformation is explained in Section 2.3.

1. Resources

1.1 ICD10

ICD10 is the most frequently used classification of diseases. The first volume contains the hierarchy of the classification system, organised in 5 levels, ranging from the 22 chapters which group together diseases according to major categories (e.g. cardiovascular diseases and neoplasms) to optional fifth character subdivision which is mainly left for national purposes (however WHO itself defined some). Each category has a name and may have (local) coding rules, comments, definitions, synonyms and dagger-asterisk cross-references.

Since the aim of ICD is to classify all possible diseases, there are a lot of “other” (e.g. J10.8 “Influenza with other manifestations, influenza virus identified”) and “not otherwise specified” (e.g. J12.9 “Viral pneumonia, unspecified”) categories in the coding system, where the formal representation of “other” is a difficult task: it is not certain that such a category is defined only by the exclusion of its siblings from its explicit parent category, but other categories with similar (clinical) meaning located in different part of the hierarchy.

1.2. DOLCE

The DOLCE is a descriptive upper-level ontology especially designed for ontology cleaning and interoperability. It incorporates a set of achievements from philosophical conceptual analysis in the past century. DOLCE takes a clear descriptive stance and avoids any revisionism: it adopts a neutral attitude towards traditional “metaphysical” questions, instead of resolving such issues it leaves open all possibilities. It also incorporates results from genuine research into mereology done by the developers themselves. DOLite+ is a modular implementation of DOLCE in OWL DL.

Extended Descriptions and Situations (EDNS) is intended to provide a framework for representing contexts, methods, norms, theories, situations, and models at first-order, thus allowing a partial specification of those entities [7]. Since DOLCE is an ontology of particulars, i.e. a system of concepts about individual entities, concepts themselves would be second-order constructions in DOLCE. This would raise theoretical and practical difficulties, therefore the designers have decided to reify these second order constructions, i.e. to represent them as first order social objects. The main concept of this reified sub-ontology is *edns:social-object*, its descendants include:

- *edns:description* is a social object which represents a conceptualization (e.g. a mental object), hence it is generically dependent on some agent and communicable. It uses concepts or figures and can be satisfied by situations.

- *edns:situation*: whilst *edns:description* is the set of reifications of theories of the second order properties of first-order entities, it is set of the reifications of these properties themselves.
 - *edns:role* is a social object that classifies endurants, as used in some description. Thus it is the descriptive counterpart of endurants.
 - *edns:course* is the descriptive counterpart of perdurants (e.g. events, states).
 - *edns:parameter* is descriptive counterpart of regions.
- EDNS is available in OWL as an extension to DOLite+.

1.3. Foundational Model of Anatomy

For the representation of anatomical entities we have selected the Foundational Model of Anatomy [8], a general purpose anatomical reference ontology. It is perhaps the most detailed ontology of the human anatomy. The representation of the ontology is frame based [9], and it contains 70 thousand concepts, consequently it is not very useable for reasoning. For that reason in a previous project [10] we have transformed it to a compact core ontology (aligned to DOLCE) represented in OWL, without losing too much information. This reduction was done by the elimination of left and right entities (e.g. instead of *Left hand*, *Right hand* and *Hand*, only *Hand* is left, with an attribute of being *mirrorImaged*), and by creating building blocks of compositional entities (e.g. we only state that *Synovial joint* must contain an *Articular capsule*, and entities like *Articular capsule of incudomalleolar joint* have been removed).

1.4. SNOMED International

SNOMED International is a combinatorial medical conceptual system covering the whole area of medicine [11]. It contains 11 modules, the most important for us are the Morphology, Function and Procedure modules. The modules typically contain primitive concepts (like tonsil and inflammation), while complex concepts can be represented by combination of simple ones (e.g. tonsillitis = tonsil + inflammation). The modules have hierarchic structure, which is reflected in the code. However there is no clear semantics of the broader-narrower relations. We decided to use SNOMED International in stead of SNOMED CT, because we only need a list of atomic building blocks for representation of functions, morphology and procedures.

- In certain cases (e.g. diabetes, autism) it is easier to represent the disease as an impairment of a certain function in stead of describing its location as a specific anatomical entity. For this purpose we use the Function module.
- For the representation of morphological entities (e.g. inflammation, carcinoma) we have selected the Morphology module. Except for the classification of neoplasms (ICD-O), it is expected that only a few concepts (e.g. infarction, necrosis, fracture) will be used.
- For the representation of procedures only very few concepts are needed (e.g. anaesthesia, transplantation). Nevertheless, we will use the Procedure module, because it enables us the compatibility with the future ICPM ontology.

We use these modules only as a plain list of terms, without the taxonomical relations. After finishing the description of the disorders based on these concepts, they will be rearranged to reference ontologies of the given domains. Functions and morphological entities will be classified as *edns:description*, while procedures as *edns:activity* (with a related concept below *edns:description*, if necessary).

1.5. Other resources

For the representation of organisms (virus, bacterium, etc.), we use the biological taxonomy. All living organism (Bacteria and Eucaria) are classified as *edns:agentive-physical-object*, while viruses as *edns:non-agentive-physical object*. The cause of the distinction is that living organisms represent their environment, while viruses not.

The classification of chemical objects is based on the International Union of Pure and Applied Chemistry nomenclature (IUPAC) [12], and represented as *common:chemical-object*. Organic compounds are classified according to their functional groups (e.g. alcohol, amine). Certain categories below Y40-Y59 "Drugs, medicaments and biological substances causing adverse effects in therapeutic use" refer not to chemicals, but to the medical effect (e.g. Y40 "Systemic antibiotics"), for this purpose we will use the ATC Classification System [13], inserted below *edns:role*.

2. The structure of the ICD10 ontology

The goal is to represent formally the meaning of each ICD category, using (if possible) already existing (bio)medical ontologies and terminologies. The ontology describing ICD10 will only contain the concepts representing each ICD category (labelled by the ICD code), and the formal definition of the category. The other information (coding rules, references, etc.) gives only additional information to the user of the coding system, therefore it is stored separately from the formal representation. However this additional information may partially determine the meaning of a category (e.g. diabetes mellitus is listed in chapter IV, except for diabetes in pregnancy in chapter XV), consequently, it is taken into account during the modelling. The hierarchical relations of ICD are also not represented in the ontology, since they may not always overlap with the hierarchy inferred from the formal definition (e.g. C09 is "Malignant neoplasm of tonsil", and C09.0 is "Tonsillar fossa", while tonsillar fossa is not a part of the tonsil).

We decided to represent ICD10 in two levels:

- First we identify the atomic disease concepts (meaningful to clinicians) behind the ICD categories. These concepts include not only diseases, but also symptoms, abnormal findings and social/environmental situations.
- Then the ICD categories are described by using these atomic disease concepts.

Both levels are represented in OWL DL, harmonised with DOLCE, because we think that a formal top level ontology acts as a conceptual framework aiding us to create a sound ontology, useful for formal reasoning. Moreover, - at least in principle - it aids ontology merging with other formal ontologies of the domain.

2.1. Representation of diseases

We defined the concept *ICD_Disease:Disorder* (*edns:description*), which subsumes diseases, syndromes, abnormal findings, signs and symptoms. If possible, we give a patho-physiological definition of such disorders, as a combination of:

- anatomical location (*edns:non-agentive-physical-object*) of the disease
- morphology (represented as *edns:description*, potentially referring to a pathological process)

- cause of the disease (e.g. infection, poisoning, procedures, mechanical forces), represented either as an *edns:description*. The causative agent (if applicable) is referred through an *edns:role*.
- Supplemental information, like acquired or congenital, or stage of the disease is modelled as an *edns:parameter*

This modelling principle is similar to that of the creators of EDNS [14].

In certain cases (e.g. diabetes) it is easier to represent the disease as an impairment of a certain function, rather than that of a certain anatomical structure. In such cases the function is represented as an *edns:description*, the impairment is a special concept of morphology. In certain cases (e.g. F80.4 "Childhood autism") the condition is really an ill-defined syndrome. In such cases these concepts are simply listed as children of a defined general category (e.g. impairments in behaviour and social interactions).

While diseases typically refer to well defined anatomical structures, symptoms typically refer to regions (e.g. abdomen). Abnormal findings are the results of measurements of physiological parameters, which can be practically represented as (an abnormal value) of an *edns:parameter* pertaining to a physiological process, represented as an *edns:action*.

2.2. Representation of ICD categories

ICD categories are modelled as *edns:description*, labelled by the ICD code. The related *edns:situation* is not represented (since it can be represented as a compositional concept). Most categories are defined by containing a disease group (e.g. J12.0 "Adenoviral pneumonia"), however are frequently extended in several ways:

- level of confidence (e.g. H40.0 "Suspicion of glaucoma"), modelled as *edns:parameter* which takes values from a *dol:abstract-region*.
- ground of diagnosis (e.g. A15 "Respiratory tuberculosis, bacteriologically and histologically confirmed"), modelled as an *edns:course*.
- cause of condition, which can be:
 - o other disease or ICD category (referred to through an *edns:role*)
 - o infection, which can be extended by mode of transmission (e.g. A82.1 "Urban rabies"), modelled as an *edns:description*.
 - o external causes, like accidents, which are modelled as *dol:event*, which typically also describe the participants as *edns:role* (e.g. V01 "Pedestrian injured in collision with pedal cycle")

The "not elsewhere classified" diseases (e.g. J12 "Other viral pneumonia") are represented as logical exclusion of the "elsewhere classified" ICD categories from the appropriate parent concepts.

2.3. Transformation of the GALEN-based representation

In a previous project the first two chapters of ICD10 has been represented using GALEN as reference ontology. The definitions of these categories will be also transformed to the new model. Since in that case we created a one-level model (i.e. directly defined the ICD categories with pathological concepts), first we will have to identify the corresponding disease concepts, and than take over the pathological definition. Because the concepts in FMA, GALEN, and SNOMED are labelled with English terms, in most cases the corresponding concepts in FMA, SNOMED, and organism ontology can be automatically identified. In certain cases the GALEN

ontology itself defined the referred disease (e.g. *Meningitis*, *Hepatitis*), in such cases we will manually add the pathological definition. After that the logical expressions between ICD categories can be automatically converted.

3. Discussion and conclusion

Contrary to the common belief that ICD10 categories classify diseases, the reality is that ICD10 classifies various other conditions, like syndromes, abnormal signs, symptoms and findings, gravity etc. Moreover, we have decided to deviate from the conceptual structure of the GALEN-based representation, because in reality the meaning of an ICD category is not a disorder, but an entity that classifies the disorders.

The taxonomy of ICD10 reflects a long history of epidemiological considerations that are different from principles of modern ontology engineering. While the historical structure can remain untouched, the formal representation of the meaning of the categories should follow these principles enabling automated reasoning and semantic interoperability in various situations.

References

- [1] International Statistical Classification of Diseases and Health Related Problems, Who, Geneva, 1992
- [2] International Classification of Procedures in Medicine, Who, Geneva, 1978
- [3] Masolo C, Borgo S, Gangemi A, Guarino N, Oltramari, A, Schneider L: WonderWeb Deliverable D17. The WonderWeb Library of Foundational Ontologies and the DOLCE ontology, preliminary Report (ver. 2.0, 15-08-2002)
- [4] Information materials about OWL can be found at <http://www.w3.org/2004/owl>
- [5] Héja G, Surján G, Lukácsy G, Pallinger G, Gergely M: GALEN based formal representation of ICD10, *Stud Health Technol Inform.* 2005;116:707-12.
- [6] Rector AL, Solomon WD, Nowlan WA, Rush TW: A Terminology Server for Medical Language and Medical Information Systems, IMIA WG6, Geneva, May 1994
- [7] Gangemi A, Mika P: Understanding the Semantic Web through Descriptions and Situations, International Conference on Ontologies, Databases and Applications of Semantics (ODBASE 2003), Catania, (Italy), November 3-7, 2003
- [8] Rosse C, Mejino JLV Jr.: A Reference Ontology for Biomedical Informatics: the Foundational Model of Anatomy. *J Biomed Inform.* 2003 Dec; 36(6): 478-500
- [9] Minsky ML: A framework for representing knowledge, in Winston P H (ed.) *The Psychology of Computer Vision* 211-277 McGraw-Hill, New York, 1975
- [10] Héja G, Pallinger G, Surján G: Restructuring the Foundational Model of Anatomy, *Stud Health Technol Inform.* 2006;124:755-60.
- [11] Coté RA, Rothwell DJ et al. (eds.): SNOMED International, College of American Pathologists, Northfield Il. USA 1993
- [12] Information about IUPAC can be found at: http://www.iupac.org/dhtml_home.html
- [13] Information about ATC can be found at: <http://www.whocc.no/atcddd/>
- [14] Gangemi A, Catenacci M, Battaglia M: Inflammation Ontology Design Pattern: an Exercise in Building a Core Biomedical Ontology with Descriptions and Situations, in Pisanelli DM (ed.): *Ontologies in Medicine*, IOS Press, Amsterdam, Berlin, 2004, pp. 64-80

A Version Management System for SNOMED CT

Josef INGENERF^{a1}, Thomas BEISIEGEL^a
Institute of Medical Informatics, University of Luebeck

Abstract. The parallel use of different versions of classifications and terminologies complicates the comparability of coded patient data. Annual adaptations of classifications like ICD-10 can be dealt with by using simple conversion tables. Terminologies like SNOMED CT have a much more complicate structure and bi-annual versions in different languages are offered in at least four representation formats. A version management system, called Terminology Version (TV) Manager, has been developed for searching and navigating in synchronized presentations of selected versions of SNOMED CT. Several functions for qualitative and quantitative comparisons of subtrees of interest are extended by accessing external software systems.

Keywords. Terminology, algorithms, versions and formats of SNOMED CT

Introduction

The parallel use of different versions of classifications and terminologies complicates the comparability of coded patient data. Annual adaptations of classifications like ICD-10 can be dealt with by using simple conversion tables. Most often recent versions are more differentiated than new ones so that an automatic mapping is possible directed to older versions. The following example is taken from the German adaptation ICD-10-GM:

Example. Mappings from the conversion table “ICD-10-GM 2006 ↔ ICD-10-GM 2007”

T23.2 ↔ T23.20 (*bidirectional automatic mapping*)

T23.2 ← T23.21 (*unidirectional automatic mapping*)

T23.2 Burn of second degree of wrist and hand

T23.20 Burn of degree 2a of wrist and hand

or Burn of unspecified second degree of wrist and hand

T23.21 Burn of degree 2b of wrist and hand

¹ Corresponding Author: Josef Ingenerf, Institute of Medical Informatics, University of Luebeck, Ratzeburger Allee 160, 23538 Luebeck, Germany; Email: ingenerf@imi.uni-luebeck.de

Although ICD-10 based code conversion is not so easy [1] there are reasons why simple conversion tables are insufficient for treating different versions of SNOMED CT:

- **Structural and quantitative complexity:** Compared to classifications like ICD-10 with about 12.000 classes organized as a simple tree the reference terminology SNOMED CT contains more than 350.000 concepts organized in poly-hierarchies and using non-hierarchical relations in concept definitions.
- **Post-coordination:** Non-hierarchical relations are used in creating composite concepts not available in SNOMED CT as delivered with pre-coordinated concepts.
- **Terminology is software:** Contrary to classifications terminologies like SNOMED CT cannot be printed. This makes it difficult for users to isolate relevant parts and identifying changes in different versions. Using (versions of) SNOMED CT is only possible by using suitable software systems.

For end users like clinicians the complexity of SNOMED CT has to be hidden in application software, e.g., by offering pre-coded picking lists for data entry. However, terminological experts creating translations, subsets, extensions or mappings to classifications as well as clinicians doing evaluation work need suitable software systems. This is certainly true when treating different versions of SNOMED CT. Several browsers are available providing more or less suitable functions for searching and navigating the huge concept system, i.e., public domain systems like CliniClue [2] and SNOB [3], commercial systems from Apelon, Inc. [4] and Health Language, Inc. [5] as well as Web-based browsers [6]. With all these systems the user can interact just with one version of SNOMED CT at once. Although a history mechanism is available for tracking e.g., outdated concepts by using relationships like “Replaced By”, the structural consequences of changes are hardly evident to the user [7]. In order to study the changes of different versions simultaneously the Terminology Version (TV) Manager has been developed. First, a short introduction to SNOMED CT and its versions is given in section 1. After that, the TV-Manger is presented in section 2. Amongst others, the functionality, architecture, extended functions by linkage to external tools, and a small walk through is provided. The results are summarized in section 3, followed by a discussion.

1. SNOMED CT and its different versions

SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms) is a comprehensive clinical terminology for clinical documentation and reporting [8]. It can be used to code, retrieve, and analyze clinical data in much greater detail compared to classifications like ICD-10. The terminology is comprised of approximately 350.000 concepts organized in 19 categories (e.g., clinical finding, procedure, and body structure), 700.000 descriptors and more than one million relationships organizing the concepts in polyhierarchies. Concepts, descriptors and relationships are identified by unique persistent identifiers. Based on defined concepts using description logics the subsumption hierarchy is inferred automatically. Table 1 shows a fully defined concept with inferred superordinate concepts that is already included (pre-coordinated) with identifier 397825006 in SNOMED CT.

Table 1. Fully defined (pre-coordinated) concept “Gastric Ulcer”

<p>CONCEPT: 397825006 Gastric ulcer (Disorder)</p> <p>DESCRIPTORS (TERMS): Gastric ulcer (1777426014, preferred), Stomach ulcer (1785985013), ...</p> <p>DEFINITION: Fully defined by ... - <i>Is a</i> 64572001 Disease (Disorder) - Group - <i>Associated morphology</i> 56208002 Ulcer (morphologic abnormality) - <i>Finding site</i> 69695003 Stomach structure (body structure)</p> <p>SUPERORDINATE CONCEPTS (inferred by a classifier) - 29384001 Disorder of stomach (disorder) - 40845000 Gastrointestinal ulcer (disorder)</p>
--

Post-coordinated concepts can be built analogously by refining value restrictions (e.g., “Pyloric antrum”) and adding relations (e.g., the sanctioned relation “Causative agent” and the qualifier “Severity”) together with value restrictions, e.g., “severe” and “Helicobacter pylori”. With that, they also can be classified, e.g., as a subconcept of the already existing pre-coordinated concept “89662003 Helicobacter-associated pyloric ulcer”. Compared to fully defined concepts primitive concepts are not defined at all (e.g., “69695003 Stomach structure *Is a* 49596003 Digestive organ structure”) or definitional criteria are necessary but not sufficient, e.g., “34000006 Crohn's disease *Is a* 24526004 Inflammatory bowel disease AND (*Associated morphology* 23583003 inflammation, *Finding site* 113276009 Intestinal structure)”. With primitive concepts (majority in SNOMED CT) the exploitation of a classifier is limited, e.g., not all inflammations at the intestinal structure are Crohn’s diseases.

SNOMED CT is released twice a year. Furthermore, there are several language versions, e.g., in Spanish. Changes to concepts and descriptions are indicated by status codes, i.e., whether they are in active use and, if not, indicating the reason for withdrawal from current use. Active concepts are marked mainly with the status codes “current” and “limited”. Inactive concepts are subdivided with using status codes like “outdated”, “duplicate” or “ambiguous”. These are included in all versions to support migration of already coded patient data. Each change in the status is recorded in a history table providing amongst others the reason for change and relationships like “Same as”, “Replaced by”, “Maybe a”, “Was a”, “Moved to” or “Moved from”. SNOMED CT is delivered in different formats.

- Native core tables:
Original flat tables with information to concepts, descriptions and relationships from IHTSDO [9]. Subsets, extensions or mappings are not regarded here.
- CLUE–Browser format:
A compiled data format used by CliniClue [10]. For usage by the TV-Manger the terminology has to be extracted by accessing the COM/OLE-API.
- UMLS-based format:
This format is based on a completely different data model [11].
- Description Logic (DL) format:
Internally, SNOMED CT is represented in description logic format called KRSS, respectively Ontylog [9, 12].

In order to be able to compare these different heterogeneous representation formats they are transformed to the distribution format as used with the native core tables [9]. Especially the DL version of SNOMED CT makes it necessary to generate a comparable concept system with respect to parent concepts and inherited criteria. We have used CEL (A polynomial-time Classifier for the description logic EL+) as a classifier that is efficient enough to classify the whole concept system, i.e., to create the isa-hierarchy automatically [13]. EL+ is the name of a tractable description language (L) employed by SNOMED CT, mainly using conjunction and existential restriction (E) as concept constructors.

2. Terminology Version (TV) - Manager applied to SNOMED CT

Based on a contract with IHTSDO (International Health Terminology Standard Development Organisation) the resources of SNOMED CT in table 2 have been obtained:

Table 2. Versions of SNOMED CT in different formats

	07.2004	10.2004	01.2005	07.2005	01.2006	07.2006	01.2007
CORE							X
CLUE		X (Ger)				X (Eng)	
UMLS	X		X			X	
DL			X				

Oriented to the core table format a common data model is defined. Import routines realizing diverse syntactic transformations are challenged by the complicate data structures underlying the UMLS Metathesaurus on the one hand and by the creation of a compliant concept hierarchy for the DL version on the other hand. In order to optimize the performance of the TV-Manager several metrics are pre-computed. For each concept the number of direct and all descendant and ancestor concepts with respect the transitive closure is determined, distinguishing amongst others types (primitive, fully defined) and status (current, outdated, ...) of concepts. The following features are provided to the user:

- Access to the selected ontologies:
 - o Filter for selecting concepts of interest with respect to concept type and status
 - o Textsearch (fulltext search or phrase search with/without soundex)
 - o Navigation in the hierarchy and via links in concept definitions
 - o Determining super- and sub-concepts of post-coordinated concepts by incremental classification with a linkage to the CEL classifier
- Structural comparison:
 - o Synchronized display of the hierarchies of all selected ontologies with the selected concept in focus
 - o With concept in focus: if available show information and hierarchical contexts in all selected ontologies; if not, show positions of the missing concept
 - o With concept in focus: List of differences with respect to similar concepts in each subtree, e.g., changes, additions and deletions in definitions

- o With concept in focus: Comparison of selected subtrees with the PROMPT-plugin in Protégé and visualization with another plugin (OWL-VIZ).
- Quantitative comparison:
 - o Display of the number of subconcepts as suffix to each concept's name.
 - o With concept in focus: Calculation of numbers of concepts in each subtree and their differences; distinguished by types and status of concepts.
 - o For the whole ontologies: Calculation of numeric metrics, e.g., differentiating concept type (primitive, fully defined) and/or status (current, outdated, ...), used relationships, leave nodes or average/longest path-length.

The TV-Manager is implemented in Java 6.0 and uses MySQL 5.0 for storing and querying the versions of SNOMED CT. This allows the TV-Manager to be platform independent being able to interact with Protégé [15] (e.g., for visualizing selected subtrees by using suitable plugins) under Windows and Unix and with CEL [14] for efficient classifying new composed concepts of interest (incremental classification) that is only available for Unix machines. In order to take advantage of both systems the selected concept definitions in core table format have to be transformed syntactically in the mentioned dialects. This is especially true for the PROMPT plugin available with Protégé that is a well known tool box for managing multiple ontologies [16]. Selecting suitable small fragments of SNOMED CT the TV-Manager provides an interface to PROMPT [17].

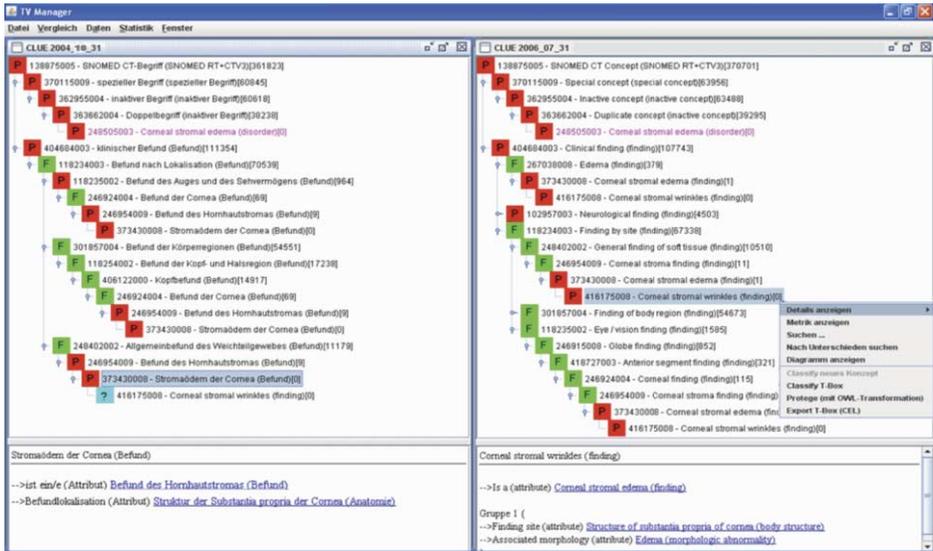


Figure 1. Comparison of two SNOMED CT versions with TV-Manager

Due to limited space just one screenshot is presented. In figure 1 the versions CLUE 10/2004 and 07/2006 are compared after the concept “Corneal stromal edema” has been searched. With this concept in focus synchronized hierarchical contexts are presented. The subconcept “Corneal stromal wrinkles” in CLUE 07/2006 is not available in CLUE 10/2004, indicated by a “?”. Inactive concepts are presented in purple colour.

3. Discussion

TV-Manger has been helpful in comparing two or more versions of SNOMED CT. The synchronized presentation of the hierarchical context of selected concepts helps to understand changes in definitions. As a side effect, this supports the study of different language versions presented next to each other. Especially finding and procedure concepts like the one in figure 1 are located very often in poly-hierarchies with many different branches to the top concept. Even with the TV-Manager it is still difficult to perform a comprehensive and efficient study of the area of interest. Until now the TV-Manager is mainly a proof of concept used by some interested terminology experts. The basic idea of comparing different SNOMED CT versions simultaneously together with making external functionality available in one software system is convincing. However, for a broader use the system the performance of the system need to be improved. Furthermore, it must be explored more carefully to what extend the four representation formats mentioned earlier can be transformed to comparable normal forms at all. One of the plans for future developments is to include the history table for being able to traverse inactive concepts in one version to active concepts in another version and vice versa. Furthermore a module for determining and visualizing consequences of new versions for already coded patient data. Simple textual changes of a concept name should be distinguished from severe changes, e.g., with modified contexts and mappings to ICD-10.

References (URLs last accessed February 15, 2008)

- [1] Schulz S, Zaiß A, Brunner R, Spinner D, Klar R (1998). *Conversion problems concerning automated mapping from ICD-10 to ICD-9*. *Methods Inf Med.* 37 (3): 254-259.
- [2] CliniClue Software, see <http://www.cliniclue.com/software>.
- [3] SNOB - A SNOMED Browser, see <http://snob.eggbird.eu/>.
- [4] Apelon, Inc.: DTS - Distributed Terminology System, see <http://www.apelon.com/products/dts.htm>.
- [5] Health Language, Inc. providing Terminology Software, see <http://www.healthlanguage.com/>.
- [6] Veterinary Terminology Services (Web-based SNOMED CT-Browser), see <http://snomed.vetmed.vt.edu/>.
- [7] SNOMED Clinical Terms® - Technical Reference Guide, January 2008 International Release (for UMLS Licensees included in the SNOMED CT package), see <http://umlsks.nlm.nih.gov/>.
- [8] SNOMED Clinical Terms® - User Guide, January 2007 Release, see <http://www.ihtsdo.org/>
- [9] SNOMED Clinical Terms® - Abstract logical models and Representational forms, January 2006 CMWG Revision (Version 5), see <http://www.ihtsdo.org/our-standards/technical-documents/>.
- [10] CliniClue Data, see <http://www.cliniclue.com/cluedata>.
- [11] SNOMED CT included in UMLS, see http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html.
- [12] Spackman KA, Dionne R, Mays E, Weis J (2002). *Role grouping as an extension to the description logic of ONTYLOG, motivated by concept modeling in SNOMED*. *Proc. of the AMIA 2002 Annual Symp.*, 712-716.
- [13] SNOMED Clinical Terms® - Transforming Expressions to Normal Forms, August 2006 Revision (Version 5), see <http://www.ihtsdo.org/our-standards/technical-documents/>.
- [14] A polynomial-time Classifier for the description logic EL⁺, see <http://lat.inf.tu-dresden.de/systems/cel/>.
- [15] The Protégé Ontology Editor and Knowledge Acquisition System, see <http://protege.stanford.edu/>.
- [16] Noy NF, Musen MA (2003). *The PROMPT suite: interactive tools for ontology merging and mapping*. *Int. J. Hum.-Comput. Stud.* 59 (6): 983-1024.
- [17] Beisiegel T (2007). *Struktureller Vergleich und Visualisierung verschiedener Versionen der Referenzterminologie SNOMED CT*. Bachelorarbeit am Institut für Medizinische Informatik an der Universität zu Lübeck.

Representation of disorders of the newborn infant by SNOMED CT®

Andrew G. JAMES MBChB MBI FRACP FRCPC^{a,1},

Kent A. SPACKMAN MD PhD^b

Departments of^aPaediatrics, University of Toronto, Toronto, Ontario, Canada and

^bMedical Informatics and Clinical Epidemiology, OHSU, Portland, Oregon, USA.

Abstract. SNOMED CT® is the most sophisticated reference terminology currently available for the representation of healthcare. An unforeseen consequence of the opportunistic evolutionary process for SNOMED CT may be that some terms for disorders of specialised clinical domains are not represented within the terminology. The SNOMED CT July 2006 release was systematically examined using the CliniClue® terminology browser to determine whether 434 terms for disorders of the newborn infant are represented within the terminology. There was complete representation for 90.8% of the terms for disorders of the newborn infant, partial representation for 6.4% of the terms, and no representation for 2.8% of the terms. Complete representation is achieved with a single, pre-coordinated SNOMED expression for 96.2% of the terms for disorders of the newborn infant that have complete representation within SNOMED CT. Nearly ninety percent of the SNOMED CT concepts that completely represent these terms have the current Concept Status but less than 40% of these concepts are fully defined SNOMED concepts. Nearly 50% of these SNOMED CT concepts have one or more synonyms. SNOMED CT® provides structured representation for the majority of this set of terms that are used for disorders of the newborn infant.

Keywords. SNOMED CT®, SNOMED Clinical Terms, Systematised Nomenclature of Medicine Clinical Terms®, clinical terminology, knowledge representation, disorders of newborn infant.

Introduction

SNOMED CT®, the Systematised Nomenclature of Medicine Clinical Terms® (SCT), is a comprehensive, concept-based, clinical terminology with a semantic model based on description logic that uniquely identifies and describes clinically relevant concepts. SCT has a polyhierarchical structure with multiple parent-child relationships together with relationships between concepts in different subtype hierarchies that define the meaning of a SNOMED concept relative to other SNOMED concepts. SCT contains more than 370,000 active concepts, more than one million active descriptions, and nearly 1.5 million relationships. The coded information for the terminology consists of

¹ Corresponding Author: Department of Paediatrics, The Hospital for Sick Children, 555 University Ave., Toronto, Ontario M5G 1X8, Canada; E-mail: andrew.james@sickkids.ca

SNOMED expressions that are references to SNOMED concepts. Clinical terms may be represented as a single, pre-coordinated SNOMED expression or a post-coordinated SNOMED compositional expression of two or more SNOMED concepts.

SCT is the most sophisticated reference terminology currently available for the representation of healthcare. The terminology has evolved from the Systematised Nomenclature of Pathology [SNOP, 1965] into the Systematised Nomenclature of Medicine [SNOMED, 1974], a reference terminology [SNOMED RT, 2001] and a clinical terminology [SNOMED CT, 2002] over a period of forty years. The content of SCT has been determined largely by the voluntary contributions of many diverse clinical groups. An unforeseen consequence of this opportunistic evolutionary process may be that some unique and clinically relevant terms of specialised clinical domains are not represented in SCT.

This paper describes the representation of a set of terms for disorders of the newborn infant by SNOMED concepts.

1. Methods

The SNOMED CT[®] July 2006 release [1] was systematically interrogated to determine if 434 terms that are used for disorders of the newborn infant are represented by SCT concepts. The terms were extracted from the Clinical Information Management System (CIMS)² used in the Neonatal Intensive Care Unit at the Hospital for Sick Children, Toronto, Ontario, Canada. The CIMS was implemented in 1999 and has been customized steadily with regular in-house upgrades in response to user defined needs. Structured data entry for disorders of the newborn infant [diagnosis] was added to the CIMS in 2003 for clinical, research and administrative purposes. Representative examples of frequently and less frequently CIMS terms for disorders of the newborn infant are provided in Table 1.

The CliniClue 2006 Terminology Browser [version 2006.2.8; October 7, 2006] [2], a look-up engine for SCT concepts, was used to systematically interrogate SCT to determine the representation of the 434 CIMS terms by SNOMED concepts. No restrictions were applied to the search. If the initial search did not return a match, alternative but semantically identical words or phrases were entered as the search term. The fully specified name, Concept Id, Description Id, Concept Status [current, limited], the status of the Definition [primitive, fully defined], the hierarchy tag, and the number of synonyms were recorded for the SCT concepts that represented the CIMS terms for disorders of the newborn infant.

The representation of each CIMS term within SCT was classified as complete, partial or absent. Complete representation was defined as the presence of a SNOMED concept for the CIMS term. The CIMS terms that were completely represented within SCT were categorized as a simple or a compositional match to the SCT concept. A simple match requires a single, pre-coordinated SCT expression that is an exact lexical match for the CIMS term. A compositional match requires a post-coordinated SCT compositional expression with two or more SCT concepts that is semantically equivalent of the CIMS term. Partial representation was defined as the presence of a related SCT concept for the CIMS term. The CIMS terms that were partially represented by a SCT concept were further examined to determine if a post-coordinated

² Sunrise Critical Care (Emtek), Eclipsys

SCT compositional expression could be used to achieve complete representation of the CIMS term. Absent representation was defined as the absence of both a SCT concept and a related SCT concept for the CIMS term.

Table 1. Representative CIMS terms for neonatal disorders

Frequently used CIMS terms	Less frequently used CIMS terms
Transient tachypnoea of the newborn, meconium aspiration syndrome, respiratory distress syndrome, bronchopulmonary dysplasia	Pulmonary interstitial emphysema, subglottic stenosis, congenital diaphragmatic hernia, pulmonary hypoplasia, pulmonary agenesis
Ventricular septal defect, patent ductus arteriosus, transposition of the great arteries, pulmonary stenosis, coarctation of aorta	Atrioventricular canal defect, hypoplastic left heart syndrome, interrupted aortic arch, total anomalous pulmonary venous drainage, congenital heart block
Gastroesophageal reflux, necrotising enterocolitis, duodenal atresia, tracheo-esophageal atresia/fistula	Meconium ileus, gastroschisis, malrotation, Hirschsprung's disease, anorectal malformation
Urinary tract infection, hypospadias, renal failure	Renal agenesis, epispadias, extrophy of the bladder
Anaemia, haemolytic disease of the newborn	Congenital thrombocytopenia
Hypoglycaemia, hypocalcaemia, hypothyroidism	Congenital adrenal hyperplasia, phenylketonuria
Hypoxic-ischaemic encephalopathy, intraventricular haemorrhage, periventricular leucomalacia, retinopathy of prematurity, seizures	Anencephaly, encephalocele, holoprosencephaly, cerebellar haemorrhage, transient myasthenia gravis, transverse venous sinus thrombosis
Congenital dislocation of the hip, talipes equinovarus	Submucous cleft palate, septic arthritis
Group B streptococcal infection, RSV infection	Congenital rubella syndrome
Intrauterine growth retardation, cephalohaematoma, Erb's palsy, non-immune fetal hydrops	Neonatal drug withdrawal, subarachnoid haemorrhage, skull fracture due to birth injury

The CIMS terms for neonatal disorders were categorised as congenital or acquired. The terms for acquired disorders were further categorised as a unique neonatal or general term. A CIMS term was identified *a priori* as a unique neonatal term if the term included the word neonatal or usage of the term was restricted to the neonatal period, and *post-priori* if the relevant SCT concept included the word neonatal.

The data for all 434 CIMS terms was analysed with simple descriptive statistics. The representation of the CIMS terms by SCT concepts is reported as the percentage of instances that meet the definition for complete, partial or absent representation.

2. Results

Four hundred and twenty-two of the 434 CIMS terms (97.2%) are represented by SCT concepts. There is complete representation for 394 CIMS terms (90.8%) and partial representation for 28 CIMS terms (6.4%). Twelve CIMS terms (2.8%) for disorder of the newborn are not represented by SCT concepts (Table 2). All 99 terms categorised as congenital are represented by SCT concepts. There is representation for 97.8% of 232 terms categorised as unique neonatal and 98% of 103 terms categorized as general.

Table 2. CIMS terms not represented within SNOMED CT

CIMS term	Category
Birth injury to peripheral nervous system, unspecified	Unique neonatal
Craniofacial	General
Hemolytic diseases of fetus and newborn, unspecified	Unique neonatal
Non-hemolytic jaundice	General
Other neonatal stroke, unspecified	Unique neonatal
Perinatal depression	Unique neonatal
Suboptimal vascular access	General
Transient disorders of calcium and magnesium metabolism	General
Unique perinatal diagnosis	Unique neonatal
Unspecified disturbance of potassium balance of newborn	Unique neonatal
Unspecified disturbance of sodium balance of newborn	Unique neonatal
Unspecified intrauterine growth restriction	Unique neonatal

Three hundred and forty-one of the 394 SCT concepts (86.5%) that completely represented CIMS terms have the current Concept Status. Complete representation is achieved with a single, pre-coordinated SNOMED expression for 379 of these 394 CIMS terms (96.2%). Fifteen of these 394 CIMS terms (3.8%) require a post-coordinated SNOMED compositional expression for complete representation within SCT. Ten of these compositional expressions require the use of the unapproved attributes stage or grade. One hundred and forty-four of the 394 SCT concepts (36.5%) are fully defined concepts. One hundred and seventy-nine of these SCT concepts (45.4%) have one or more synonyms.

3. Discussion

SCT is the most sophisticated clinical terminology currently available for the representation of health care. The terminology fulfills the desired requirements for a clinical terminology because the thirteen desiderata for a clinical terminology of the twenty-first century [3,4] have been integrated into the design of SCT. The additional SNOMED URU criteria [understandability, reproducibility, usefulness] together with the use of natural terms, component persistence and history tracking have also enhanced the performance of SCT.

Many studies [5-14] have confirmed that the SNOMED family of terminologies provides a more comprehensive and consistent representation for clinical care than other classifications and terminologies. The SNOMED family provides very good representation for general medical [5-12] and the specialised medical domains of cardiovascular disease and ophthalmology [13,14]. The SNOMED CT[®] July 2006 release provides very good representation for the CIMS terms that represent the disorders of the newborn infant. More than 97% of the terms are represented by SCT concepts although 10% of the terms are only partially represented. Twelve CIMS terms (2.8%) are not represented by SCT concepts. Six of these unrepresented terms are legacy terms inherited from the International Classification of Diseases that contain the word *unspecified*. These terms should not exist within SCT since their inclusion would contravene the desiderata for a clinical terminology of the twenty-first century [4].

More than 95% of the CIMS terms for a disorder of the newborn infant are represented by a pre-coordinated SCT expression. The 15 CIMS terms that required a post-coordinated SNOMED compositional expression for complete representation are terms that demand greater granularity than those that do not require post-coordination. Ten of the compositional expressions use the unapproved attributes of stage or grade to achieve complete representation of the severity of necrotizing enterocolitis and retinopathy of prematurity. The documentation that less than five percent of 394 completely represented CIMS terms require post-coordinated SNOMED compositional expression supports the view that SCT is a rich, comprehensive and granular clinical terminology.

All concepts within a clinical terminology should be fully defined to enable the computational process of auto-classification and the computation of equivalence and subsumption between different methods of expressing the same meaning. Nearly 40% of the completely represented CIMS terms are represented by fully defined SCT concepts. Many SCT concepts are not fully defined at the present time because they are very general or non-clinical concepts, or the attributes or values necessary for the full definition of the concept are missing. Furthermore, it takes time to fully define over 370,000 current concepts in a large terminology that has over one million descriptions and nearly one and a half million relationships. The attainment of fully defined status for more than 70% of SCT concepts is one of the stated goals for SNOMED. The representation of disorders of the newborn infant cannot be considered to be complete until the majority of the relevant SNOMED concepts are fully defined. Nearly 50% of the SCT concepts that completely represent CIMS terms have synonyms. The existence of one or more synonyms enhances the terminology's functionality as an interface terminology, and therefore, should facilitate the accurate entry of valid data.

Although neonatology is a specialised domain of medicine with many disorders that are unique to the neonatal period, the domain represents the health care for a specific, well-defined population as distinct from other specialties that represent health care for those with disorders related to a specific organ system. The majority of the terms for disorders of the newborn infant have been derived from the domains of general medicine and surgery. The representation of CIMS terms for disorders of the newborn infant by SCT concepts exceeds 90% and is consistent with the representation by SCT for other domains that have their origins within general medicine and surgery [5-14].

4. Conclusion

SNOMED CT[®] provides structured representation based on description logic for the majority of this set of terms that are used for disorders of the newborn infant. An increase in the number of fully defined SCT concepts and the number of concepts with one or more synonyms will enhance the terminology's computational processes and the interface functionality of SCT for clinicians and other users. Future activities should focus on increasing the number of fully defined SCT concepts and synonyms for relevant SCT concepts for disorders of the newborn infant.

References

- [1] SNOMED Clinical Terms® Core Content for the July 2006 Release. Northfield, IL, USA; 2006. Available at www.snomed.org [accessed October 27, 2006].
- [2] CliniClue 2006 [version 2006.2.8, October 7, 2006]. Clinical Terminology Services from The Clinical Information Consultancy. Available at www.cliniclue.com [accessed November 6, 2006].
- [3] Chute CG, Cohn SP, Campbell JR, for the ANSI Healthcare Informatics Standards Board Vocabulary Working Group and the Computer-based Patient Records Institute Working Group on Codes and Standards. A framework for comprehensive health terminology systems in the United States: Development guidelines, criteria for selection, and public policy implications. *J Am Med Inform Assoc.* 1998; 5:503-510.
- [4] Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods Inf Med.* 1998; 37:394-403.
- [5] Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR. The content coverage of clinical classifications. For The Computer-Based Patient Record Institute's Work Group on Codes & Structures. *J Am Med Inform Assoc.* 1996; 3:224-33.
- [6] Campbell JR, Carpenter P, Sneiderman C, Cohn S, Chute CG, Warren J. Phase II evaluation of clinical coding schemes: completeness, taxonomy, mapping, definitions, and clarity. CPRI Work Group on Codes and Structures. *J Am Med Inform Assoc.* 1997; 4:238-51.
- [7] Humphreys BL, McCray AT, Chen ML. Evaluating the coverage of controlled health data terminologies: report on the results of the NLM/AHCPR large scale vocabulary test. *J Am Med Inform Assoc.* 1997; 4:484-500.
- [8] Dykes PC, Currie LM, Cimino JJ. Adequacy of evolving national standardized terminologies for interdisciplinary coded concepts in an automated clinical pathway. *J Biomed Inform.* 2003; 36:313-25.
- [9] Wasserman H, Wang J. An applied evaluation of SNOMED CT as a clinical vocabulary for the computerized diagnosis and problem list. *AMIA Annu Symp Proc.* 2003:699-703.
- [10] Penz JF, Brown SH, Carter JS, Nguyen VN, Sims SA, Lincoln MJ. Evaluation of SNOMED coverage of Veterans Health Administration terms. *Medinfo 2004*; 11:540-4.
- [11] Elkin PL, Brown SH, Husser CS, Bauer BA, Wahner-Roedler D, Rosenbloom ST, Speroff T. Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists. *Mayo Clin Proc.* 2006; 81:741-8.
- [12] McClay JC, Campbell J. Improved coding of the primary reason for visit to the emergency department using SNOMED. *Proc AMIA Symp.* 2002:499-503.
- [13] Min Z, Baofen D, Weeber M, van Ginneken AM. Mapping OpenSDE domain models to SNOMED CT. Applied to the domain of cardiovascular disease. *Methods Inf Med.* 2006; 45:4-9.
- [14] Chiang MF, Casper DS, Cimino JJ, Starren J. Representation of ophthalmology concepts by electronic systems: adequacy of controlled medical terminologies. *Ophthalmology.* 2005; 112:175-83.

Exploratory Analysis of Medical Coding Practices: the Relevance of Reported Data Quality in Obstetrics-Gynaecology

Diana LUNGEANU^{a1}, Daniela ZAHARIE^b, Stefan HOLBAN^c, Elena BERNAD^d,
Maria BARI^c, Rodica NOAGHIU^a

^a*Department of Medical Informatics, University of Medicine and Pharmacy of Timisoara, Romania*

^b*Faculty of Mathematics and Computer Science, West University of Timisoara, Romania*

^c*Faculty of Computer Science, Polytechnic University of Timisoara, Romania*

^d*Department of Obstetrics-Gynaecology, University of Medicine and Pharmacy of Timisoara, Romania*

^e*Department of Obstetrics-Gynaecology, University of Medicine and Pharmacy of Bucharest, Romania*

Abstract. We aimed at identifying a suitable data analysis approach to investigate potential patterns in the current medical coding in obstetrics and perinatal care. We processed the data reported for 2006 in DRG files from three Romanian university clinics of obstetrics-gynaecology and found substantial differences in the coding practices. Based on the evidence we found with a poor usage of the coding instruments, we concluded that using objective methods and quantifiable measures in analyzing the medical coding could help putting things into the right perspective and bring support for the need for formal education of medical record administrators and coders where such programmes do not exist, e.g. in Romania.

Keywords. Data analysis tools, coding, diagnosis related, outcomes research and measurement

Introduction

The *Diagnosis Related Groups* (DRG) system was initially developed as an information management tool to monitor quality and use of services but is now widely used as a prospective payment system in most European countries [1, 2]. Although comprising data collected mainly at the discharge moment and typically used for measuring the severity of illness, so resulting in the case-mix index for each specific hospital as a basis for financing and/or reimbursement, the reported DRG files also contain valuable information for quality management within the hospital sector [3, 4]:

¹ Corresponding Author: Diana Lungeanu, Department of Medical Informatics, University of Medicine and Pharmacy, P-ta E. Murgu 2, 300041 Timisoara, Romania; E-mail: dlungeanu@umft.ro

main and secondary diagnoses (the latter important as co-morbidities) based on the *International Classification of Diseases v10* or ICD-10 [5, 6], procedures' codes, length of stay, etc.

As in 2005 the DRG system became compulsory for the hospital care in Romania, we tried to use this reported minimum data set to perform a patient profiling in obstetrical and perinatal care. However, in processing the data we encountered quality and reliability problems (data initially collected from a single hospital), so we decided to further investigate the medical coding patterns, as data validity and reliability are important criteria in choosing appropriate sources and measurements for assessing healthcare services [2]. The DRG system is at the beginning in Romania and the code assignment affects the case-mix and the financing in this prospective payment system. It should thus not be surprising that the hospitals try to "accommodate" the code usage to their perceived needs for financing, trying to balance this trade-off against the correct coding from a medical point of view. This has been a sensitive problem wherever the system was introduced [1, 7].

Taking this into account, we decided to perform an analysis of the coding practices in obstetrics and evaluate the use of codes, aiming at discovering the factors which influence the current practices.

1. Methods Used in Investigating Medical Coding Practices

Our goal was to investigate the coding results in hospitals of obstetrics-gynaecology in order to determine whether there were any coding patterns or "common practices" (i.e. good vs. bad habits). Empirical data comprising the ICD-10 codes used for the principal and secondary diagnoses (i.e. the main medical problem and the comorbidities or complications) were imported from the DRG application compulsorily used in Romanian hospitals for reimbursement purposes.

We combined basic statistical analysis for the available ICD-10 codes' usage with data mining techniques. We used *SPSS v.15* and *Weka 3-4* to process the data.

1.1. Statistics Used to Characterize the Medical Coding Practice

The ICD-10 utilization statistics consisted of: (i) number of secondary codes (reflecting co-morbidities) recorded in the DRG files; (ii) percentage of codes used, of those specific for each specialty (obstetrics-gynaecology and perinatal care, respectively); (iii) number of codes covering up to 50% of cases for the principal and secondary diagnoses; (iv) a concentration index borrowed from industrial economics and introduced to healthcare by Spangler et al [4].

The degree of concentration across the coding categories was measured using the Hirschman-Herfindahl index (HHI), defined as it follows in Eq. (1).

$$\begin{aligned} & \text{of } n \text{ diagnosis codes that have } p_i \text{ (} i=1,2,\dots,n \text{) percentage} \\ & \text{shares for a specialty in a hospital,} \\ & HHI = \sum_{i=1}^n p_i^2 \end{aligned} \quad (1)$$

The value range for HHI is between 0 and 10000: the larger the number, the higher the degree of codes' concentration (i.e. a small number of recorded/reported codes account

for a disproportionately large number of all medical diagnoses or problems, comorbidities, etc. in that specialty and/or hospital). An HHI number below 1000 is considered to indicate a fair degree of concentration for medical codes' usage [4].

1.2. Clustering Methods Used to Explore the Medical Coding Practice

The aim of a data clustering task is to identify homogeneous groups of similar data which satisfy at least the following conditions: (i) data in each group are sufficiently similar; (ii) data in different groups are sufficiently dissimilar. Our investigation started from a concern regarding the validity and reliability of the reported data set, so we tried to see whether the observed cases were "parceled" across the available ICD-10 codes, i.e. we were looking for patterns of diagnostic codes' utilization. We used the actual codes (as nominal variables) and the length of stay (as a numerical variable, illustrating the severity of the episode of care), expecting to find natural clusters comprising similar clinical cases, on the condition that a reliable and consistent manner of coding was used with respect to the World Health Organization's recommendations [5, 6]. Moreover, we assumed that for equivalent teaching hospitals with similar pathologies one can expect to obtain similar clustering results. We tried the *K-means* and *Expectation Maximization* (EM) methods.

The *K-means* algorithm [8] is a partitioning approach based on the idea of minimizing the average distance between the data in a cluster and the cluster's center (i.e. the prototype). It consists of an iterative process which starts from some randomly selected centers, one for each cluster to be discovered. It can be easily adapted to categorical data by replacing the average of data in a cluster with the element which has a maximal similarity with all the other elements of the cluster (the so-called *medoid*), so we considered it as appropriate for our data set with both nominal and numerical variables.

In contrast with *K-means* where the clusters are disjoint, the *EM* algorithm proposed by Dempster [9] relies on a statistical interpretation of the clustering problems: data are considered as a mixture of probability distributions and the technique seeks to identify this mixture's parameters. Similarly to *K-means*, *EM* is also based on an iterative process involving two main steps at each iteration: (i) *expectation step*, which computes the probabilities of data with respect to the current values of the mixture parameters; (ii) *maximization step*, which estimates the new parameters of the probability distributions corresponding to clusters, based on a maximum likelihood approach. This algorithm allowed us a statistical approach in determining the optimum number of clusters for each data set: (a) we first applied the *EM* algorithm on each dataset in order to estimate the appropriate numbers of clusters; (b) those values were further used to compute the clusters' prototypes by the *K-means* technique.

2. Results and Discussion

We analyzed data for 2006 in DRG files from three university hospitals, one in a distant geographic location from the other two: HB (6862 cases), HC (5831 cases), HO (10301 cases). In this paper, we focus on the exploratory methods for evaluating ICD-10 codes' usage. We also tried to get a perspective on the medical coding in Romanian hospitals and to identify the potential patterns in the current medical coding practices.

Table 1 presents the statistical analysis. When calculating the codes' usage statistics for new-born babies (NB), we considered as specific the 408 codes from classes P00-P96 (*Certain conditions originating in the perinatal period*) and Z30-Z39 (*Persons encountering health services in circumstances related to reproduction*). Similarly, when calculating the codes' usage statistics for OG, we considered as specific the 715 codes from classes C51-C58 (*Malignant neoplasms – Female genital organs*), D06-D07 (*In situ neoplasms – Carcinoma in situ*), D25-D28 (*Benign neoplasms*), D39 (*Neoplasms of uncertain or unknown behaviour*), N70-N77 (*Inflammatory diseases of female pelvic organs*), N80-N98 (*Noninflammatory disorders of female genital tract*), O00-O99 (*Pregnancy, childbirth and the puerperium*), and Z30-Z39 (*Persons encountering health services in circumstances related to reproduction*).

At first, when comparing medical coding patterns, we might think it would be difficult to compare HHI for different specialties for they had completely different characteristics (e.g. intensive care compared to radiology or hematology). However, as the statistics prove in Table 1, even in the specific case of one single specialty (i.e. obstetrics-gynaecology or neonatal care) there is a high degree of heterogeneity across the three hospitals, as well as within the same hospital for the two different specialties.

Table 1. Statistics regarding the ICD-10 codes' usage in the three hospitals (HB, HC, HO) for the wards of new-born babies (NB) and obstetrics-gynaecology (OG). HHI stands for Hirschman-Herfindahl index.

Characteristic	HB		HC		HO	
	NB	OG	NB	OG	NB	OG
No of cases	2197	4665	2390	3441	2854	7447
Length of stay (in days)						
min—max	0—31	0—34	0—369	0—367	0—45	0—80
mean	4.65	4.83	9.39	6.32	5.22	6.38
(std dev)	(2.20)	(3.26)	(32.21)	(15.44)	(2.11)	(5.71)
No of secondary codes/case						
min—max	0—8	0—10	1—18	0—20	0—14	0—15
mean	0.69	1.66	3.69	11.32	2.61	4.90
(std dev)	(0.73)	(2.53)	(1.47)	(6.07)	(2.54)	(1.58)
* % of ICD-10 codes used for principal diagnosis	24%	24%	11%	7%	**28%	33%
* No of principal codes covering 50% of cases	8	3	7	2	**4	3
* HHI for principal diagnosis codes	609	1293	1500	2195	**1009	1275
No of secondary codes covering 50% of cases	3	1	2	9	3	4
HHI for secondary codes	966	3113	1853	407	1357	1100

* Codes from the classes A00-B99 (*Certain infectious and parasitic diseases*), D50-D89 (*Diseases of the blood and blood-forming organs and certain disorder involving the immune mechanism*), Q00-Q99 (*Congenital malformations, deformations and chromosomal abnormalities*), and R00-R99 (*Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified*) were not included among those specific for female genital, maternity care or perinatal period, therefore not included in the denominator of any calculated ratios and percents. When encountered in the files, they had very low frequencies, so we considered them as outliers and completely excluded them from the codes' usage statistics (i.e. included neither in the numerator nor the denominator).

** One hundred six cases had as principal diagnosis D18.0 (*Haemangioma, any site* from the class of *Benign neoplasms*) and were excluded from the statistics regarding the ICD-10 usage for new-born babies

Although all three clinics are of similar size and of the same specialty, significant differences were obtained with respect to the number of clusters found by the *EM* method: 8 (HB), 16 (HC), and 6(HO). Table 2 presents the prototypes we identified by the *K-means* method for one of the hospitals. The codes that appear in a cluster's prototype (consisting of the mean values for the numerical attribute and the mode for the categorical ones) could be interpreted as best-describing the instances (i.e. cases) composing that group. Looking at the codes' pattern for HB, clusters 2, 3, and 6 are identical, the only difference being in the length of stay. We chose to use length of stay as a clustering attribute, for it is one of the criteria influencing the DRG codes and the case-mix. Surprisingly the length of stay is not reflected in the coding pattern, though the longer the stay, the more complex should have been the case and that extra-burden of the disease should have been reflected in the secondary diagnoses. This flatness or lack of detail might be due to a rather poor coding, which is also suggested by the HHI for OG (Table 1 - high values for both the principal and the secondary codes). On the other hand, a good aspect is reflected by cluster 0, with D25.0 as the principal diagnosis code, correctly corresponding to the patients with the main diagnosis in the *Neoplasms* class (C00-D48). However, there still is a problem with this cluster, as the secondary codes are related to postpartum examination/care and to the babies' care, so conflicting with the principal code. It also seems the code Z24.6 (specific to the new born babies) "engulfed" all the secondary codes for the OG clusters (for which the HHI is high as well).

Table2. Clusters for HB identified by the *K-means* algorithm (8 clusters, 13 attributes, 6832 instances). For each cluster, the specialty, the number of instances and the percent are specified. Although all the available secondary diagnoses were considered in the clustering process, only the first four are presented in the table.

Cluster centroids	Length of stay (m±sd)	Princ diag	Sec diag1	Sec diag2	Sec diag3	Sec diag4
Cluster 0 OG 446 (6%)	9.58±4.92	D25.0	Z39.2	Z24.6	Z24.6	Z24.6
Cluster 1 OG 1025 (15%)	6.82±2.69	O82.0	Z39.2	Z24.6	Z24.6	Z24.6
Cluster 2 OG 1136 (17%)	4.35±2.69	O80.0	Z39.2	Z24.6	Z24.6	Z24.6
Cluster 3 OG 1754 (26%)	2.28±0.78	O80.0	Z39.2	Z24.6	Z24.6	Z24.6
Cluster 4 NB 1037 (15%)	5.75±2.50	P59.9	Z39.2	Z24.6	Z24.6	Z24.6
Cluster 5 NB 207 (3%)	4.73±1.65	P21.1	Z24.6	Z23.2	P59.9	P59.9
Cluster 6 OG 304 (4%)	7.65±2.39	O80.0	Z39.2	Z24.6	Z24.6	Z24.6
Cluster 7 NB 953 (14%)	3.44±0.99	Z38.0	Z39.2	Z24.6	Z24.6	Z24.6

The titles (i.e. medical meaning) for the ICD-10 codes appearing in Table 2 and/or in the text are: D25.0 *Benign neoplasm – Submucous leiomyoma of uterus*; O80.0 *Spontaneous vertex delivery*; O82.0 *Single delivery by elective caesarian section*; P21.1 *Mild and moderate birth asphyxia*; P59.9 *Neonatal jaundice (physiological, intense, prolonged), unspecified*; Z24.6 *Need for immunization against viral hepatitis*; Z38.0 *Singleton, born in hospital*; Z39.2 *Postpartum care and examination - Routine postpartum follow-up*

The concrete results for the other two hospitals were completely different from HB, i.e. different patterns corresponding to the clusters. On the other hand, they were similar in the sense of their intriguing clusters' centroids, in a consistent manner with the HHI values (i.e. the higher HHI values, the meaningless the clusters).

3. Conclusions

Using exploratory analysis techniques, we investigated the medical coding in three university hospitals that offer top-quality medical care. We employed basic statistics regarding the specific codes' usage, a concentration index borrowed from economics, and clustering techniques from data mining.

The conclusions are consistent with the initial perspective: there is a poor usage of the medical coding instruments. Based on these results and on the fact that two of the hospitals were located in the same geographic area (i.e. with similar patient profiles), we are confident the heterogeneities and difficulties are mainly due to the inexperience in applying the constraints of the DRG system. However, poor data quality leads to unreliable performance indicators and prevents appropriate and timely decisions, entailing subsequent costs. Therefore, we do think that using objective methods and quantifiable measures in analyzing the medical coding practices could help putting things into the right perspective and eventually obtaining good quality medical data.

Our study underpins the need for formal education programmes for medical record administrators (e.g. university programmes) to prepare professionals in charge with supervising medical coding activities in hospitals. In Romania we do not have such professionals and, moreover, most often the coders are the medical doctors themselves.

Acknowledgements

We appreciate the valuable contribution of Dr. Dan Navolan from the University Hospital of Obstetrics-Gynaecology "Dr. D. Popescu" from Timisoara. We also kindly thank Professors Ioana Moisil and Bernard Richards for their encouragement and useful comments on the early results of our work.

The study was supported by grant 99-II CEEX 03 - INFOSOC 4091/31.07.2006 from the Romanian Ministry of Education and Research.

References

- [1] Schreyogg J, Statgardt T, Tiemann O, Busse R. Methods to determine reimbursement rates for diagnosis related groups (DRG): a comparison of nine European countries. *Health Care Manage Sci* 2006; 9: 215-223.
- [2] Kane RL. *Understanding health care outcomes research* (2nd Edition). Sudbury MA: Jones and Bartlett Publishers, 2006.
- [3] Calle JE, Saturno PJ, Parra P, Rodenas J, Perez MJ, San Eustaquio F, Aguinaga E. Quality of the information contained in the minimum basic data set: results from an evaluation in eight hospitals. *European Journal of Epidemiology* 2000; 16: 1073-1080.
- [4] Spangler WE, May JH, Strum DP, Vargus LG. A data mining approach to characterizing medical code usage pattern. *Journal of Medical Systems* 2002; 26(3): 255-275.
- [5] WHO. ICD-10 Online. <http://www.who.int/classifications/apps/icd/icd10online> (Last access Oct 2nd 2007)
- [6] WHO. ICD-10. International statistical classification of diseases and related health problems. 2nd Edition, 10th Revision 2004. http://www.who.int/classifications/icd/ICD-10_2nd_ed_volume2.pdf (Last access Oct 2nd 2007)
- [7] Stausberg J, Lang H, Obertacke U, Rauhut F. Classification in routine use: lessons from ICD-9 and ICPM in surgical practice. *Journal of AMIA* 2001; 8(1): 92-100.
- [8] Kanungo T, Mount DM, Netanyahu NS, Piatko CD, Silverman R, Wu AY. An efficient k-means clustering algorithm: analysis and implementation. *IEEE Trans on Pattern Analysis and Machine Intelligence* 2002; 24(7): 881-892.
- [9] Dempster A, Laird N, Rubin D. Maximum likelihood estimation from incomplete data via the EM algorithm. *J Royal Stat Soc* 1977; 39(1): 1-38.

Use Of Super-Concepts To Customize Electronic Medical Records Data Display

P. MASSARI^a, S. PEREIRA^a, B. THIRION^a, A. DERVILLE^b, SJ. DARMONI^a

^a *CISMeF, Rouen.University Hospital & GCSIS, LITIS EA 4108, Institute of Biomedical Research, University of Rouen, France*

^b *IS@S Company. Paris. France*

Abstract. Patient medical record systems (MRS) merely offer static applications, in which mostly unstructured text is linked to coded data. In these applications the more common presentation is a time oriented one, which does not allow easily for data and information retrieval. Concept oriented views based on super-concepts (metaterms) initially defined in CISMeF to optimize Web medical search, was implemented in our MRS as specialties views. This work shows that these terminological tools are able to facilitate information retrieval.

Keywords. Medical Records Systems, Computerized, Data Display, Information Storage and Retrieval, Abstracting and Indexing

Introduction

The architecture of patient medical record systems often respect the European standard HISA [1]. In these systems, Electronic Health Records (EHR) are structured to a certain degree, according to patient record system used, and most of the EHR contain a majority of unstructured texts [2], these texts being linked to contacts¹ and procedures² as described in HISA [1]. These entities, contacts and procedures are also associated with respective ICD10 and procedure codes.

The most common presentation of these data in the MRS is a time related presentation, referring the dates and periods of care. In this model information retrieval can be difficult when patients have a long medical history. To optimize information retrieval in EHR, the problem-oriented medical record was introduced as a concept in 1968 [3]. It does not yet have a wide spread because of the requirement on structuring patient-data entries [4].

Another way is to implement appropriate views, using terminological tools. The CISMeF metaterms [5] are super-concepts which were initially defined to draw together MeSH terms from the MeSH thesaurus [6]. CISMeF's metaterms correspond to medical specialties (e.g. cardiology), types of medical procedures (e.g. surgery) or

¹ HISA definition of contact : event during which the clinical state of a subject of care is under the active consideration of healthcare agent.

² Procedures included office visits.

health topics (e.g. diagnosis, therapy). Creating metaterm (or super-concepts) came up initially to optimize information retrieval in the CISMeF catalogue, in particular to maximize the recall [7]. We have decided to reuse the super-concepts in a different environment (MRS) using different health terminologies. ICD 10 for coding diagnosis and several health procedure classifications in use in France : CCAM [8] (since September 2005) and CDAM (before 2005) for therapeutic and diagnosis procedures and ADICAP for pathological exams. CISMEF metaterms embedded in the EHR could offer to clinicians customized views of the EHR elements, that could be more efficient than the classical time-oriented presentation.

The aim of this study is to describe and evaluate individual medical records sorted by typology of elements and by medical specialties based on terminological tools, using CISMeF super-concepts.

1. Material and Methods

1.1. Patient medical record system of the Rouen University Hospital

EHR was introduced in 1992 in the Rouen University Hospital [9] and took into account medical contacts, laboratory results, discharge ICD codes, medical procedures codes, medical procedures reports and discharged medical reports. Between 1992 and the end of 1999, the medical record system was implemented on a mainframe system, while since the end of 1999 a new medical record system working on an Oracle data base has been in use, allowing to take into account the same data and specific structured data. Currently EHR includes all the computerized data from 1992 until now, corresponding to 1.2 million patient records, 8.6 million contacts, 9.34 million medical procedures, 1.2 million discharge reports and 3.1 million medical procedure reports.

1.2. CISMeF metaterms and MeSH

CISMeF ([French] acronym for Catalog and Index of French Language Health Resources on the Internet) [1] is a quality-controlled health gateway. It was designed to catalog and index the most important and quality-controlled sources of institutional health information in French in order to allow end-users to search them quickly and precisely (N=36,589). Its URLs are <http://www.chu-rouen.fr/cismef> or <http://www.cismef.org>. CISMeF uses two standard tools for organizing information : the MeSH thesaurus [6] from the US National Library of Medicine and several metadata element sets, in particular the Dublin Core metadata format (URL:<http://www.dublincore.org>) [10]. However, the MeSH thesaurus was originally intended to index scientific articles for the Index Medicus and for the MEDLINE database. In order to customize it to the broader field of health Internet resources, we have been developing several enhancements [1] to the MeSH thesaurus, with the introduction of two new concepts, respectively metaterms (MT) and resource types (RT). CISMeF RT are an extension of the publication types of MEDLINE. Creating metaterms were firstly designed to optimize information retrieval in CISMeF and cope with the relatively restrictive nature of these medical specialties as MeSH (keywords) descriptors. The MeSH thesaurus does not allow to have a global vision of a medical specialty. Therefore, in the CISMeF terminology, metaterms can be considered as

“meta-concepts”. Metaterms have been manually selected by the chief medical librarian (BT). The semantic links between metaterms and MeSH terms, MeSH subheadings and CISMef resource types are based on his know-how and expertise of medical specialists of the RUH. There is a 0 to N relations between CISMef metaterms and MeSH terms, MeSH subheadings and CISMef resource types. In September 2007, the number of metaterms in the CISMef terminology was 123. The comprehensive list of metaterms is available at the following URL : http://doccismef.chu-rouen.fr/liste_des_meta_termes_anglais.html.

1.3. CISMef metaterms and ICD10, CCAM, CDAM and ADICAP classifications

A clinician (PM) and the chief medical librarian (BT) reviewed the list of CISMef metaterms to be reused in this context (N= 66 out of 123, 54 %). Semantic links were manually created between super-concepts and each code of ICD 10, CCAM, CDAM, ADICAP classifications (see Table 1). As for MeSH thesaurus there is a 0 to N relations between CISMef metaterms and codes of ICD10, CCAM, CDAM, ADICAP classifications. As an example, the metaterm Cardiology is semantically linked with the ICD 10 code I50.0-Congestive Heart Failure, the CCAM code "DZQM006-Échographie-doppler transthoracique du cœur et des gros vaisseaux" (Echocardiography, transthoracic) and the ADICAP code "BHCZ-Biopsie endomyocardique" (Endomyocardial biopsy).

Classification	Number of codes	Number of semantic links	Min-Max by codes
ICD10	10,505	13,650	1-3
CCAM	7,389	12,538	1-5
CDAM	7,699	13,508	1-4
ADICAP*	279	372	0-3

Table 1 : Semantic links between super-concepts and classifications in use in France

* A simplified ADICAP nomenclature is used in Rouen University Hospital

Using metaterms to make queries on clinical consultations requires the indexation of consultation service points. A partial evaluation of the manual semantic links between super-concepts and CCAM codes was performed by automated indexing tools [11].

1.4. EHR views using super-concept

To filter data of one or more medical specialties, customized views using metaterms have been implemented in our MRS. Codes and associated metaterms taken into account are : ICD10 discharge codes of hospitalization, CCAM and CDAM (before 2005) for technical medical procedures, ADICAP for pathological exams. These queries allow to make containers of data, classified by type of EHR elements (stays, diagnosis, surgical procedures...) and filtered by medical specialties.

1.5. Evaluation

The potential advantage of specialty-oriented views of EHR as compared to the "more classical" time related views, depends on the number of elements contained in the patient records. Because they have a potential interest for complex patient records with multiple reports, a preliminary evaluation needs to study the distribution of patients records size. We have studied the size of patient records taken in charge during the first trimester 2006. It was performed by cardiologists, lung specialists and gastroenterologists.

2. Results

2.1. Size of the health records

Between 1 January-31March 2006, 81,471 patients came to the Rouen University Hospital for hospitalization or consultation. 33.46% of the medical records contain more than 20 medical contacts, and 34.01% more than 20 medical procedures.

2.2. Time oriented view versus specialties views

Illustrations of time oriented and specialty views as shown on figure 2.

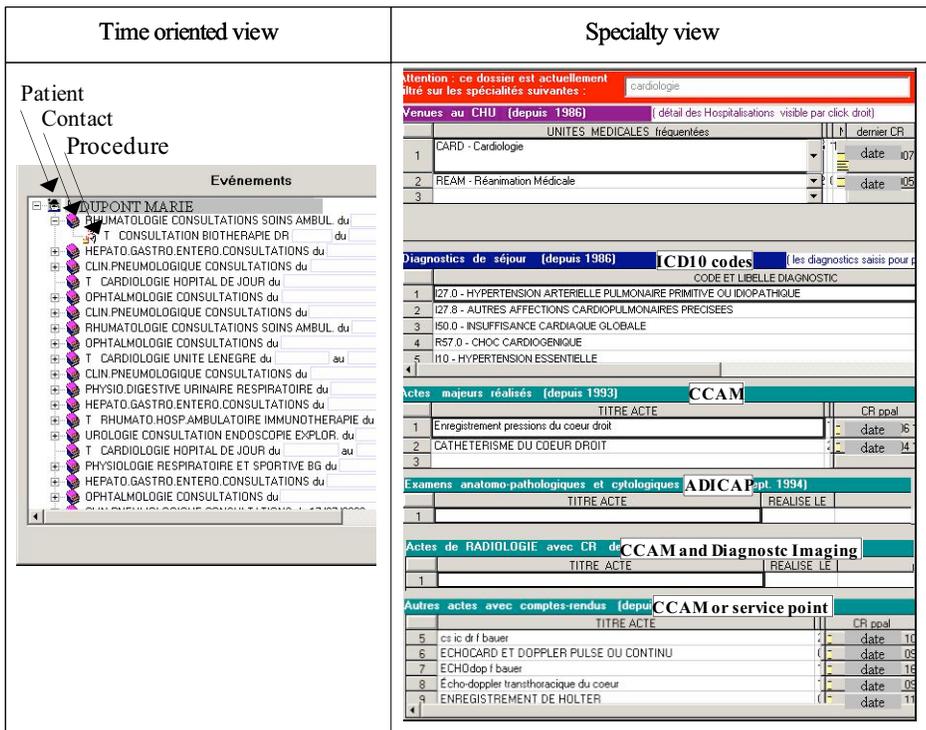


Figure 2 : Time oriented vs specialties views in Rouen EHR

The EHR used to illustrate these views is the one of a patient who has a four years medical history in RUV (see figure 2). It takes into account 62 outpatient contacts and 41 inpatient ones. The time oriented view shows in a three levels treeview including patient, contacts, and medical procedures. Linked data (codes, reports...) are accessible after selecting the element. To review the cardiological problems of this patient by the time oriented view, a clinician would have to search contacts and procedures linkable to heart disease and to read the linked data. The specialty view only displays the contacts and procedures linked with the metaterm cardiology. Seven inpatient contacts are retrieved out of 41, 5 in cardiology units, 2 in the intensive care unit (Réanimation Médicale). For these 7 contacts one or more ICD10 discharge codes are linked with the metaterm cardiology (I50.0-Congestive Heart Failure, I27.0-Primary Pulmonary Hypertension, R57.0-Cardiogenic Shock). The numbers of this EHR elements filtered by the metaterms cardiology and pneumology are shown in table 2.

Type of data	Number of elements in the whole record	Number of elements filtered by "cardiology"	Number of elements filtered by "pneumology"
Contacts	18	7	12
ICD10 discharge codes	38	6	6
Major procedures	16	2	8
Pathological reports	6	0	3
Radiological procedures	24	0	3

Table 2 : Distribution of the elements of the medical record

2.3. Comparison of the two view types of in information retrieval

A clinician in care of this patient would want to see all the cardiological informations or only the reports of a chosen procedure, for example reports of echocardiographies. In this EHR there are 3 echocardiographies, if the clinician (or the patient) do not know when these procedures were performed, without specific tools finding these 3 reports is nearly impossible. Knowing them, the user will have to select the contact, unfold the treenode and open each report, it takes less than 5 minutes to do it. Using the specialties view, no matter if the end-user knows the procedures dates or not, the three echocardiography records can be found in less than 2 minutes.

The evaluation of specialties views based on CISMef metaterms in EHR was considered as satisfying by the CISMef team and the RUH, therefore we have decided to implement this specialty view tool in the real environment of the RUH EHR .

3. Discussion

This paper presents another use of the CISMef metaterms, originally developed for the optimization of information retrieval and categorization. The CISMef metaterms are now semantically linked with different medical terminologies (ICD 10, CCAM, CDAM and ADICAP) in order to provide the clinicians customized views of EHR elements. This semantic links are manually established. Therefore, the quality of the CISMef metaterms relies on the know-how of this manual indexing. Fortunately, the

CISMeF team includes the chief medical librarian (BT) who has 12 years of experience in manual indexing using health terminologies. The fifteen years history of EHR in RUH [9] means that a significant part of patients records in the data base contains a large number of recorded events. The information and data retrieval in these records is difficult, and need specific tools [10]. Improving information retrieval can be done by different types of views, the effectiveness of concept-oriented views was reported by several authors [13] [14]. More recently, a second generation of these tools, using ontologies to define fundamental concepts, their properties, and interrelationships within a particular domain, have been described [15].

4. Conclusion

EMR display data and retrieval information can be optimized by specialties views. Super-concepts initially designed for medical Web search can be reused to create these views. These super-concepts could make possible in the future the creation of problem-oriented views, without the requirement of structuring patient data entries.

5. References

- [1] Healthcare Information System Architecture (HISA) : prENV 12967-1; CEN/ TC251, [Online] [cited 2008 Feb 22] 1997; URL : <http://www.tc251.wgiv.nhs.uk/pages/hisa.asp>
- [2] Bayegan E, Nytro O. A problem-oriented, knowledge-based patient record system. *Stud Health Technol Inform.* 2002;90:272-6.
- [3] Weed LL. Medical records that guide and teach. *The New England Journal of Medicine*, 278(12), March 21, 1968.
- [4] Lundsgaarde HP, Fischer PJ, Steele DJ. Human problems in computerized medicine, University of Kansas Publications in Anthropology Number 12, Lawrence, Kansas, 1981.
- [5] Douyère M, Soualmia LF, Névéal A, Rogozan A, Dahamna B, Leroy JP, Thirion B, Darmoni SJ: Enhancing the MeSH thesaurus to retrieve French online health resources in a quality-controlled gateway. *Health Info Libr J* 2004; 21(4):253-61.
- [6] Nelson SJ, Johnson WD, Humphreys BL: Relationships in Medical Subject Headings in Relationships in the organization of knowledge. In Bean and Green, eds. Kluwer Academic Publishers, 2001. p. 171-84.
- [7] Gehanno JF, Thirion B, Darmoni SJ. Evaluation of meta-concepts for information retrieval in a quality-controlled health gateway. *Proc AMIA Symp.* 2007:269-73
- [8] Rodrigues JM, Trombert-Paviot B, Baud R, Wagner J, Rusch P, Meusnier F. Galen-In-Use: an EU Project applied to the development of a new national coding system for surgical procedures: NCAM. *Stud Health Technol Inform.* 1997;43 Pt B:897-901.
- [9] Massari P, Smuraga I, Froment L, Boudehent S, Czernichow P, Streiff J, Baldenweck M, Hecketsweiler P. Application de gestion des dossiers informatisés du CHU de Rouen. Cinquièmes Journées Francophone d'Informatique Médicale. Genève 9-10 juin 1994.
- [10] Dekkers M, Weibel S: State of the Dublin Core Metadata Initiative. *D-Lib Magazine* April 2003: vol 9. Number 40.
- [11] Pereira S, Massari P, Joubert M, Darmoni S. Utilisation de métatermes pour la recherche d'information dans les dossiers médicaux. *JFIM* 2007, Bamako, Mali, Janvier 2007
- [12] Powsner SM, Tufté ER. Graphical summary of patient status. *Lancet* 1994; 344:386-9.
- [13] Dore L, Lavril M, Jean FC, Degoulet P. An object oriented computer-based patient record reference model. *Proc Annu Symp Comput Appl Med Care.* 1995;377-81.
- [14] Zeng Q, Cimino JJ. Evaluation of a system to identify relevant patient information and its impact on clinical information retrieval. *Proc AMIA Symp.* 1999;642-6.
- [15] Elisabeth B, Oystein N, Anders G. Ontologies for Knowledge Representation in a Computer-Based Patient Record 14th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'02). 2002. p. 114.

Semantic Web Ontology Utilization for Heart Failure Expert System Design

Marin PRCELA^a, Dragan GAMBERGER^a and Alan JOVIC^b

^a *Rudjer Boskovic Institute, Zagreb, Croatia*

^b *Faculty of Electrical Engineering and Computing, Zagreb, Croatia*

Abstract. In this work we present the usage of semantic web knowledge representation formalism in combination with general purpose reasoning for building a medical expert system. The properties of the approach have been studied on the example of the knowledge base construction for decision support tasks in the heart failure domain. The work consisted of descriptive knowledge presentation in the ontological form and its integration with the heart failure procedural knowledge. In this setting instance checking in description logic represents the main process of the expert system reasoning.

Keywords. knowledge management, knowledge bases, cardiovascular, expert systems, heart failure, reasoning

Introduction

Medical decision support systems are challenging because of the complexity and richness of the medical knowledge involved. Building a decision support system, which can make the procedures of diagnosis, prognosis and therapy more effective and reliable for the patient, and which is optimal in the use of medical and clinical resources, is yet an unattained goal and still presents a great challenge. It also imposes as a test bed for the knowledge representation formalisms which validates their adequacy and sufficiency in such applications.

It has been recognized in many cases that ontologies are appropriate for knowledge encoding within different systems [1,2,3,4,5]. Also, it has been noted that the application of general-purpose ontology reasoners is very beneficial since in that way the knowledge becomes more sharable and maintainable. The possibilities of using the reasoners in specific cases has been widely explored and tested [5,6,7]. In settings where, due to more demanding system requirements, such utilization is not adequate the system tends to become more complex and more difficult to maintain because some amount of knowledge and the reasoning procedure are encoded within the application itself [1,4,8]. Such knowledge representation approaches tend to be system-specific what scales down their reusability.

Semantic web ontology language OWL has recently emerged as de-facto standard for intelligent applications that utilize the ontologies as knowledge formalization tool. OWL, in combination with the SWRL rule language and with domain-independent reasoners, provides a generally recognized expert system development framework.

In this paper we describe the utilization of OWL in medical expert systems applications. We start by presentation of the descriptive ontology constructed for the heart failure domain and then analyse the possibility to include also procedural knowledge in the same ontological representation. Finally, based on experiments with real application we compare rule based reasoning with ontological reasoning for the procedural type of the knowledge.

1. Knowledge Representation

Modern expert systems generally recognize a few knowledge types and make a clear distinction between them by imposing distinct formalization means and distinct usage routines. Even in the philosophical domain, the knowledge is divided into *descriptive* and *procedural* knowledge.

Descriptive knowledge (also referenced as conceptual knowledge) describes the concepts in the domain, and the relations among them. In that way, every concept is described by defining its relation to other, previously defined concepts. On the other hand, *procedural* knowledge describes the procedures and actions that should be taken in given situations. In that sense, the descriptive knowledge is commonly treated as a construct for which practical usage has yet to be stated in the system while the procedural knowledge is very narrowly defined and operationalized, and clearly states what has to be done in specific situations [9,10]. The procedural knowledge commonly refers to the “know what” of the domain, while the procedural knowledge refers to the “know how” of the domain.

The third type of the knowledge recognized in the expert systems is *factual* knowledge. It refers to formalization of facts that describe the given situation, i.e. the problem that is currently being solved. Compared to other knowledge types, the relevance of the factual knowledge is restricted to the ongoing decision tasks and as such considered as generally uninteresting.

2. Heart Failure Knowledge Base

The first step in the process of building the knowledge base for the heart failure domain has been the construction of the descriptive heart failure ontology. The ontology is constructed in OWL by the Protégé tool. It is available from the project website (<http://www.heartfaid.org/links.php>).

The second step in the knowledge base development has been collection and formalization of the related procedural knowledge. This knowledge has been presented in the form of 10 sets of rules. In its development we used only terminology systemized by the previously constructed descriptive ontology.

2.1. Heart Failure Ontology

The HF ontology presents the formalized description of concepts for the heart failure domain. It includes basic HF concepts, properties that characterize patients, all relevant diagnostic examinations and tests, and treatment procedures. The ontology also includes other cardiovascular system related concepts as well as concepts related to

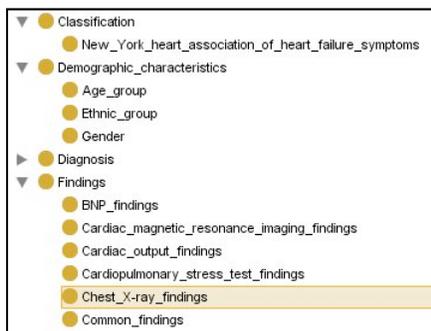


Figure 1. Excerpt from the domain description ontology representing the class hierarchy.

other organs when they are connected with HF. The information presented in the ontology has been obtained by human interpretation of guidelines for congestive and acute heart failure.

In its current form the ontology presents the detailed taxonomic overview of the HF domain with around 200 classes describing HF related concepts. These concepts are interconnected with super-class and sub-class relations into a hierarchical tree-like structure. At the basic level there are five relevant super-classes: *HF_concept*, *Patient_characteristic*, *Patient*, *Testing*, and *Treatment*. Figure 1 presents the Protégé tool displaying some of the classes from the HF ontology.

Individuals or instances are members of the classes and typically present exhaustive list of concrete concepts relevant for the class. The realized ontology includes more than 2000 individuals. When possible, classes are specified with their CUI number (Concept Unique Identifier according to UMLS) and with a list of synonyms. For example, for the class Heart diseases its CUI is C0018799 and its synonyms are *Disorder_of_heart*, *Cardiac_diseases*, *Cardiopathy*. Finally, the ontology contains properties that connect individuals in different classes. These properties are relevant because they enable introduction of relations among concepts. For example, individual *Valvular_heart_disease* from the class *Heart_valve_disease* is indicated by the individual *Dyspnea* from the class of *Signs_and_symptoms*. Or that *Hyperkalemia* from the class *Potassium_disorder* may be caused by medications like *Potassium_sparing_diuretics* or *Spironolactone*. The names of these properties are *Indicated* and *MayBeCausedByMedication*. The HF ontology includes definitions of more than 100 properties.

2.2. Heart Failure Procedural Knowledge

Production rules in a form “IF *some condition is true* THEN *make some action*” are a widely used approach for the presentation of procedural knowledge. At the knowledge presentation level it is very important that production rules can be easily understood and corrected by medical experts. In this way the major advantage of presenting procedural knowledge in the form of production rules is that they present formal enough way to present knowledge that can be used by the decision support system and that at the same time medical experts can easily control the expected performance of the system.

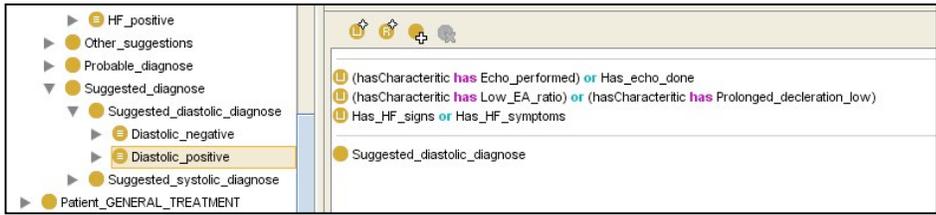


Figure 2. Example of concept constructor used for procedural knowledge encoding.

Figure 2 illustrates the possibility to present procedural knowledge in the OWL form. The OWL concept descriptors are used to formalize the conditions while the conclusions or actions that are made by the rule are represented as named OWL classes (concepts). In Figure 2 is presented the rule for the diastolic heart failure diagnosis. The presented conceptualization of the procedural knowledge is relevant because it enables its integration with before described descriptive knowledge. The additional advantage is that by the transformation from the rule form into ontological form the procedural knowledge must be ordered into a tree of sub-classes that stimulates systematization of this knowledge.

The basic task of our expert system is to check on patients characteristics and to act upon them. The basic concept of ontological procedural knowledge is the concept of *Patient*. We have assigned to that concept properties that we have found meaningful, like for example *hasCharacteristic* property which allows multiple *PatientCharacteristic* instances, or *hasTestData* which contains the instances with the numerical values of the patients test measurements. The descriptive part of the knowledge defines what possible characteristics patient might be described by. All classes representing procedural knowledge is a tree of subclasses of the class *Patient*.

3. Reasoning in OWL

The major consequence of the transformation of the procedural knowledge into ontological form and its integration with descriptive knowledge is that decision making can be completely performed by the reasoning procedures on ontologies. The experience and conclusions apply also for expert systems in other domains, particularly in other medical domains.

3.1. Reasoning in Descriptive Knowledge

One can recognize two main knowledge profiles in descriptive domain knowledge. The first is defined by the generality relations among instances and classes, as well as by the generality relations among subclasses and super-classes. In this way for each concept presented by some instance there is a series of *is-a* relation. For example, it means that *Cardiomegaly is-a Cardiac_hypertrophy* while *Cardiac hypertrophy is-a Heart disease*. The second profile of the descriptive knowledge comes from properties that define relations between individuals, such as *Indicated* or *MayBeCausedByMedication*, mentioned before.

The logic part of the OWL language (concept constructors), as we have noticed, in this case appeared to be rather superfluous. We have found that the knowledge is

substantially pre-defined and rather static, and that there was no need for describing the terms by concept constructors. The descriptive knowledge took a shape of terminology, and found a purpose just in defining a domain of discourse by listing the concepts within it and placing them in a hierarchical structure. Reasoning in the descriptive knowledge is reduced to mere propagation of *is-a* relationships down the class hierarchy, and in that way is imposed only as a structure preparation phase for reasoning in the procedural knowledge.

Here we should emphasize that this does not in any case reduce the importance of the descriptive knowledge. Domain description provides a basis for the procedural knowledge, and poor design of descriptive knowledge significantly reduces the potency of the complete system.

3.2. Reasoning in Procedural Knowledge

All reasoning tasks in the description logics are reducible to single one, e.g. satisfiability or subsumption [3]. Regarding the computational complexity, application of one or another reasoning task does not impose the additional constraints on the system. In our case, the *instance checking* takes the main role, since it assigns the patient individuals into the specific classes which represent the actions that should be performed on the patient, e.g. patient X to the class *PerformXRayTest*. Instance checking in a way simulates the execution of classical procedural rules. Due to the specific usage of the system, and due to the specific setting of the system, other reasoning tasks do not take such significant roles, although they are helpful in some situations. Namely, in the process of knowledge base building, *satisfiability check* and *consistency check* may detect some amount of contradictions in defining the concept constructors, and hence provide us with some kind of debugging tool, but such paradigm is neither requirement nor standard in classical expert systems.

By using exclusively OWL reasoning we have constrained the expressiveness of the procedural knowledge to the OWL syntax and to the reasoner semantics. In general this setting is appropriate for cases where the procedural knowledge does not require complex mathematical expressions or algorithmic control flow (like functions or loops).

The differences in reasoning between OWL and procedural rules are:

- Data transformations - Production rules generally support the common operations on data (e.g. math operators), while the description logics do not have that possibility. The cardinality restrictions are as close as the description logics have come to the numerical operations on data.
- Control flow - The common thing in procedure definition are the control flow primitives, which enable executing a statement block repeatedly or in a specific order (loops, branching, jumping, subroutines, etc.). The description logics do not use this paradigm.
- Open/Closed world semantics - Description logics use the open world semantics, which understands that the knowledge base in every moment might be incompletely defined, i.e. some statements in the knowledge base might be missing. The closed world semantics assumes that the knowledge in the knowledge base is complete. One of the crucial features of closed world assumption is negation-as-failure, which concludes that given statement is false if it is not currently reachable that it is true within the knowledge base. The production rules might follow either approach. For example SWRL semantics assume open world, while Jess semantics assume closed world.

- Reasoning tasks - Description logics perform many reasoning tasks, like *satisfiability*, *subsumption*, *classification*, *instance checking*, *consistency*, etc. Production rules have different approach; the main task is to update the knowledge base if some conditions are fulfilled.

4. Conclusion

A drawback of our approach is reasoning on data values. OWL has poor handling of numerical attributes, and therefore extension of system is necessary. The classical extension of such framework is SWRL, which is usually used to encode the procedural knowledge. In our case we have used it exclusively for simple data manipulation, e.g. determining whether the *E/A_fraction* of a patient is lower than 0,5. This has shown to be sufficient in most cases. Still, it is not recommended to use it for calculating more complex expressions, e.g. *body mass index*. This calculation is done externally (in system specific component), and loaded into the knowledge base. However, by this we have excluded some relevant amount of knowledge out of the knowledge base. Additionally this makes the factual knowledge preparation phase significantly more complex.

OWL has already demonstrated its relevance in many semantic web applications. The idea of using it as the expert system framework is not new but the originality of our approach is to use it for the conceptualization of the procedural knowledge. The major advantage is natural integration of descriptive, procedural, and factual knowledge.

References

- [1] E. Coiera. *Guide to health informatics*. (2nd ed.). Arnold, London, 2003.
- [2] W.M.Tierney, J.M. Overhage, B.Y. Takesue, L.E. Harris, M.D. Murray, D.L. Vargo, C.J. McDonald. *Computerizing guidelines to improve care and patient outcomes: the example of heart failure*. J Am Med Inform Assoc. **2** (1995), 316-322.
- [3] F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, P.F. Patel-Schneider, eds, *The description logic handbook*, Cambridge University Press, 2002, p. 5-44.
- [4] P.A. de Clercq, A. Hasman, J.A. Blom, H.H.M. Korsten, *Design and implementation of a framework to support the development of clinical guidelines*. Int J Med Inform **64** (2001), 285-318.
- [5] N.F. Noy, D.L. Rubin, M.A. Musen, *Making Biomedical Ontologies and Ontology Repositories Work*. IEEE Intell Syst **19** (2004) (6), 78-81.
- [6] D.L. Rubin, O Dameron, Y. Bashir, D. Grossman, P. Dev, M.A. Musen. *Using ontologies linked with geometric models to reason about penetrating injuries*. Artif Intell Med **37** (2006), 167 - 176
- [7] S.W. Tu, M.A. Musen, *Representation Formalisms and Computational Methods for Modeling Guideline-Based Patient Care*. Proc First Europ Workshop on Computer-based Support for Clinic Guidelines and Protocols, 2001, p.125-142.
- [8] W. Ceusters W, B. Smith, *Ontology and Medical Terminology: why Descriptions Logics are not enough*. Proceedings of the Conference Towards an Electronic Patient Record (TEPR 2003), San Antonio, 10-14 May 2003
- [9] J.R. Star, *On the Relationship Between Knowing and Doing in Procedural Learning*. In B. Fishman & S. O'Connor-Divelbiss, eds., Proceedings of the Fourth International Conference of the Learning Sciences, Mahwah, NJ: Erlbaum, 2000, p. 80-86.
- [10] M. Peleg, L.A. Gutnik, V. Snow, V.L. Patel. *Interpreting procedures from descriptive guidelines*. J Biomed Inform **39** (2006), 184-195.

Standards and Biomedical Terminologies: the CEN TC 251 and ISO TC 215 Categorial Structures. A step towards increased interoperability

Jean M. RODRIGUES^{a,1}, Anand KUMAR^a, Cédric BOUSQUET^a,
Béatrice TROMBERT^a.

^a *University of Saint Etienne, Department of Public Health and Medical Informatics,
St Etienne, France.*

Abstract: Among different biomedical terminologies standardisation strategies the European Standard Body CEN TC 251 followed by the ISO TC 215 have stated that it was not possible to convince the different European or international member states using different national languages to agree on a reference clinical terminology or to standardise a detailed language independent biomedical ontology. Since 1990, they have developed an approach named categorial structure as a step standardising only the terminologies model structure. The methodology and the review of the different existing categorial structures are presented as a step towards increased interoperability between biomedical terminologies thanks to conformity to a minimum set of ontological requirements.

Keywords: Standard; Biomedical terminology; Categorial Structure; Ontology;

Introduction

The standards and the biomedical terminologies have appeared since a long time as two different worlds. Most of developed countries maintain, update and modify their own coding systems for procedures, as well as national adaptations of ICD, in order to manage and to fund their health care delivery. The most significant efforts were done in Australia with ACHI (Australian Classification of Health Interventions) or ICD10 AM [1], in Canada with the Canadian Classification of Health Interventions (CCI) [2] developed by the Canadian Institute for Health Information and in France with CCAM (Classification Commune des Actes Médicaux) [3].

Natural language expressions show inconsistencies and ambiguities as assessed by biomedical ontology driven tools [4]. Related knowledge bases consist of multi-hierarchies of concepts organized by subsumption and associative relations. These knowledge representations are named biomedical ontology [5] [6]. The related

¹ Corresponding Author: JM Rodrigues, CHU de St Etienne, SSPIM/CIM42, Chemin de la Marandière, 42 270 Saint Priest en Jarez, France, rodrigues@univ-st-etienne.fr

automated language generation techniques include the linkage of lexicons from different national languages [7]. The most important achievements are GALEN (Generalised Architecture for Languages, Encyclopaedias and Nomenclatures in Medicine) [8], and FMA (Foundational Model of Anatomy) [9].

On the other hand, the standardisation in health informatics started in the US with the HL7 user group. The European Standard Body CEN TC 251 WG2 (Comité Européen de Normalisation Technical Committee 251 Working Group 2) and later the International Standard Organisation ISO TC 215 WG3 elaborated and developed a standard approach for biomedical terminology named Categorical structure. We address the rationale and definition of Categorical structures in part 2 and the review of the existing Categorical structures standards in part 3. We finally discuss the role of this standard approach as a compromise between one single standard for biomedical terminology and ontology based terminologies on the way to a future more complete semantic interoperability.

1. Hypothesis behind the CEN categorial structure approach

1.1. Rationale

As a prerequisite it was clearly stated at the beginning of the CEN standard process that it was not realistic to impose within Europe a standard terminology to health care professionals even as a reference terminology or a pragmatic terminology. The two main supportive arguments were that European countries speak different natural languages and that different health care professionals within the same natural language do not convey the same meaning through a particular terminology.

CEN has considered in the early 90s and ISO in the late 90s it not possible to standardise the clinical knowledge in a single biomedical ontology which could support one reference clinical terminology or different types of clinical terminologies. First the ontology developments were starting and needed consolidation and assessment of feasibility. Second the standards needed to facilitate developments in biomedical terminologies and not prevent the quickly evolving volume of terms used for different goals. Finally clinicians and medical recorders were not convinced that the formal representation of the biomedical terminologies by description logic tools which is an important point in the definition of computer ontology [10] was realistic. For these opposite challenges a third way named Categorical structure was designed and applied to some priority domains of biomedical terminologies.

1.2. Definition.

The CEN Categorical structure was defined within some linguistic variations [11-22], as a minimal set of health care domain constraints to represent a biomedical terminology in a precise health care domain with a precise goal to communicate safely. It is a definition of a minimal semantic structure describing the main properties of the different artefacts used as terminology (controlled vocabularies, nomenclatures, coding systems and classifications): a model of knowledge restricted to 1) a list of semantic categories; 2) the goal of the Categorical structure; 3) the list of semantic links between semantic categories authorised with their associated semantic categories; 4) the minimal constraints allowing the generation and the validation of well formed

terminological phrases. Any biomedical artefact claiming conformance to the standard shall attach with the data sent the Categorical structure of the terminology used. The Categorical structure shall satisfy the 4 constraints but can add more constraints.

2. Methodology of the Categorical structures

2.1. EN 1828 Categorical structure for surgical procedures [11]

We first explain precisely one Categorical structure to make explicit what is a Categorical structure and what it is not. Figure 1 shows the hierarchies of the semantic categories (left tab) and their associated semantic links (right tab).

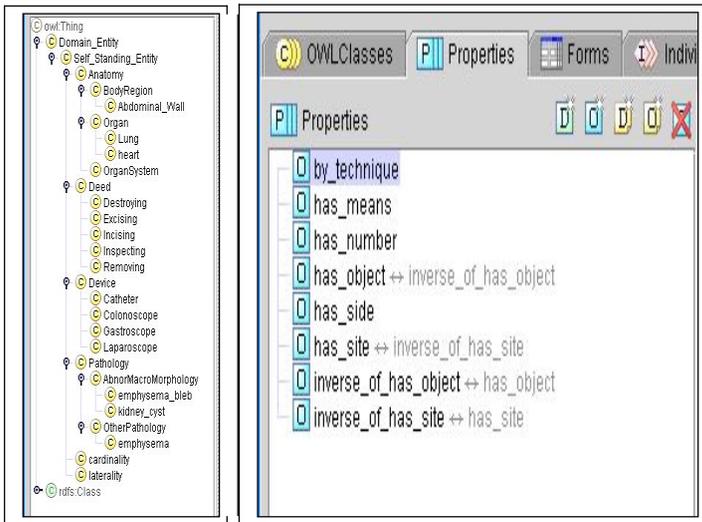


Figure 1. Semantic categories (left tab) and semantic links (right tab)

- 1 The different main semantic categories are Anatomy, Deed, Device and Pathology with 2 qualifiers categories cardinality and laterality.
- 2 The semantic links are *by_technique*, *has_means*, *has_object* (with inverse), *has_site* (with inverse), *has_side* and *has_number*.
 - 2.1 *has_object* is authorised between deed and anatomy or device or pathology
 - 2.2 *has_site* is authorised between device or pathology and anatomy
 - 2.3 *has_means* is authorised between deed and anatomy, device or pathology
 - 2.4 *by_technique* is authorised between deed and deed
- 3 The minimal constraints required
 - 3.1 A deed and *has_object* shall be present
 - 3.2 Anatomy shall always be present either with a *has_object* or with a *has_site*
 - 3.3 Use of pathology shall be restricted to macroscopic lesion and to cases where it allows to differentiate the procedure from procedures using the same deed and the same anatomy;

3.4 When *by_technique* is used the deed on the right side of the semantic link must be conform to the rules 3.1, 3.2 and 3.3.

The categorial structure allows ensuring that new terms describing surgical procedures are associated to a formal definition consistent with a common template. For example we evaluated the introduction of six new rubrics. Definitions of these six rubrics were entered in the Protégé editor [23] and classified with the Racer inference engine [24]. Concepts not consistent with the standard were identified during classification and represented as red circled concept nodes in Protégé.

2.2. The Categorial structures [11] to [16] and [18] to [22]

There are six full CEN or ISO standards for Categorial structure: CEN EN 12 264 [13] for the definition of a Categorial structure, CEN EN 1828 [11] for surgical procedures, ISO 1804 [12] for nursing terminologies, CEN EN 12 611 [14] for clinical laboratory, CEN EN 13 940 [15] for continuity of care, and CEN EN 15 521 [16] for human anatomy. On the other hand five Categorial structures are still in a provisional status on medical devices [18] and [20], medicinal products [19], vital signs [21] and patient findings and problems [22].

The standards will be revised in the five coming years by a joint process of two Working Groups from CEN and ISO. This will give the opportunity to harmonise the wording and definitions and to check their applications around the world. Finally the EN 15 521 for terminologies of human anatomy has introduced more extended constraints by prescribing detailed semantic categories and semantic links for the knowledge in the field of anatomy.

3. Results and Discussion

There are at least three different ways to address the standard issues. The first of them can be called a single standard. This is the Family of classifications used since the nineteenth century by the International Classification of Diseases (ICD) and now by the WHO Family of International Classifications (FIC). The same rationale continues to be followed when it comes to revising the ICD into its eleventh revision. However the size of knowledge is increasing so quickly with biomedical research and technology developments that maintenance and update costs are huge and never quality assured. It is an unending work based on divergent individual and domain expert opinions.

The second methodology which has emerged since 15 years can be called the biomedical co-operative approach based on formal ontology and on the biomedical ontology tools coordinated with natural language processing and web based tools [4-9]. It provides a new perspective to semantic interoperability by de-multiplying the workload organised by disseminated social computing. It is complex but separates terminology artefacts from the logic of knowledge and the linguistic characteristics of the different national languages. By associating clinical specialist domain experts, ontology experts and linguists this infrastructure gives the opportunity to assess clinical terminology by logical expertise and as well ontological representation by clinical domain experts working in their national language.

Between these two trends is a third methodology where the international standardisation organisations have the responsibility to support or assist in improving semantic interoperability. Biomedical terminologies need to conform to a minimum set

of ontological requirements. This has been the choice of the CEN European standard body within Technical Committee 251 and ISO Technical Committee 215 for terminology who has proposed the Categorical Structure. The categorical structure proposes a frame for a light ontological organisation to ensure standardisation of the knowledge representation of terminologies. The work began in the 90s with the goal to move from a syntactic or functional semantic interoperability level 1 to a semantic interoperability of level 2 where the recipient is able to understand the meaning of terms used by the sender but cannot process them as SAFELY as he can do with his own terms and meaning (semantic interoperability level 3). Developers of new terminologies or new versions of existing terminologies should conform to this standard in order to increase interoperability. It is limited but supporting the dissemination of knowledge based on a representation for any clinical and biomedical terminologies and a pedagogic step to ontology and semantic interoperability.

With the wider use of CEN categorical structure in the near future, which is imminent given the fact that European standards are mandatory for member states, would provide further insights into the advantages of using such a methodology. Future steps would include studies and trials to compare its clinical and administrative benefits to other methodologies.

Acknowledgements

We wish to thank our partners in the CEN TC251 WG2 and ISO TC 215 WG 3 and the convenors Hendrik Olesen, Ann Harding, Magnus Fogelberg and Chris Chute and terminology and biomedical ontology experts namely Alan Rector from University of Manchester and Barry Smith from the University at Buffalo. Special thanks to our partners in the GALEN program and namely Pieter Zanstra, Egbert van der Haring, Robert Baud, Jeremy Rogers, and the Stanford team for their kindness and the open quality of their website.

References

- [1] National Centre for Classification in Health see <http://www3.fhs.usyd.edu.au>
- [2] Canadian Classification of Health Interventions <http://secure.cihi.ca/cihiweb/dispPage.jsp>
- [3] Trombert-Paviot B, Rodrigues J-M, Rogers J, Baud R, van der Haring E, Rassinoux A-M, Abrial V, Clavel L, Idir H. GALEN: a third-generation terminology tool to support a multipurpose national coding system for surgical procedures. *Int J Med Inform* 2000;58-9:71-85.
- [4] Rodrigues J-M, Rector A, Zanstra P, Baud R, Innes K, Rogers J, Rassinoux A-M, Schulz S, Trombert Paviot B, ten Napel H, Clavel L, van der Haring E, Mateus C. An ontology driven collaborative development for biomedical terminologies: from the French CCAM to the Australian ICHI coding system. *Stud Health Technol Inform*. 2006;124:863-8..
- [5] Gruber T. A translation approach to portable ontology specifications. *Knowledge Acquisition* 1993; 5(2):199-220.
- [6] Smith B, Ceusters W. Towards industrial strength philosophy: how analytical ontology can help medical informatics. *Interdisciplinary Science Reviews* 2003; 28:106-11.
- [7] Baud RH, Rodrigues J-M, Wagner JC, Rassinoux A-M, Lovis C, Rush P, Trombert-Paviot B, Scherrer J-R. Validation of Concept Representation Using Natural Language Generation. In: Masys DR (Ed.). *AMIA Annu Symp Proc*. 1997: 841.
- [8] Rector A. Compositional Models of Medical Concepts: Towards Re-usable Application-Independent Medical Terminologies. In: Barahona P and Christensen JP (Eds.). *Knowledge and Decisions in Health Telematics*. Amsterdam: IOS Press, 1994; pp. 109-14.

- [9] Rosse C, Mejino JVL. 2003. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform.* 36:478-500.
- [10] Semantic Interoperability seminar. Brussels February 2005
- [11] CEN AFNOR EN NF 1828:2002. Health informatics – Categorial Structure for classifications and coding systems of surgical procedures.
- [12] EN ISO -18104, 2003 – Health informatics – Integration of a reference terminology model for nursing
- [13] CEN AFNOR EN NF 12264:2005. Health informatics – Categorial Structure for concept systems.
- [14] EN 1614: 2006, Health Informatics – Representation of dedicated kinds of property in laboratory medicine
- [15] EN 13 940: 2006, Health Informatics – System of concepts to support continuity of care.
- [16] EN 15 521:2007. Health informatics – Categorial Structure for terminologies of human anatomy
- [17] ISO 17115: 2007, Health informatics — Vocabulary for terminological systems
- [18] ENV 12 611: 1996, Medical Informatics – Categorial structure for system of concepts - medical devices
- [19] ENV 12 610:1997 Medical Informatics – Medicinal products identification
- [20] **ISO CD 15225:2000 – Nomenclature – Specification for a nomenclature system for medical devices for the purpose of regulatory data exchange**
- [21] **ENV 13734:2000 Health Informatics – Vital signs information representation**
- [22] **ISO/DTS 22789:2007, Health informatics – Conceptual framework for patient findings and problems in terminologies**
- [23] Noy NF, Fergerson RW, Musen MA. The knowledge model of Protege-2000: Combining interoperability and flexibility. 2nd International Conference on Knowledge Engineering and Knowledge Management (EKAW'2000), Juan-les-Pins, France. 2000.
- [24] Haarslev V and Möller R. Description of the RACER System and its Applications. In Proceedings International Workshop on Description Logics (DL-2001), Stanford, USA, 1.-3;131–141. 2001.

How Granularity Issues Concern Biomedical Ontology Integration

Stefan SCHULZ^{a,1}, Martin BOEKER^a and Holger STENZHORN^{a,b}

^a *Institute of Medical Biometry and Medical Informatics,
University Medical Center Freiburg, Germany*

^b *Digital Enterprise Research Institute (DERI), Galway, Ireland*

Abstract. The application of upper ontologies has been repeatedly advocated for supporting interoperability between domain ontologies in order to facilitate shared data use both within and across disciplines. We have developed BioTop as a top-domain ontology to integrate more specialized ontologies in the biomolecular and biomedical domain. In this paper, we report on concrete integration problems of this ontology with the domain-independent Basic Formal Ontology (BFO) concerning the issue of fiat and aggregated objects in the context of different granularity levels. We conclude that the third BFO level must be ignored in order not to obviate cross-granularity integration.

Keywords. Formal ontology, domain ontology

Introduction

The application of upper ontologies (also known as top-level ontologies) has been advocated for supporting interoperability between domain ontologies in order to facilitate the shared use of data both within and across disciplinary boundaries. Ideally, any upper ontology should be restricted to only a small set of domain-independent and highly general categories (types), relations, and constraints. Currently, several upper ontologies co-exist with each one differing greatly in terms of their respective user community, topical focus and logico-philosophical foundations [7].

In biomedicine, the UMLS Semantic Network [6], the GALEN High Level Ontology [9], the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [1], the General Formal Ontology (GFO) [3] as well as the Basic Formal Ontology (BFO) [2], all have been applied for this purpose. The need for bridging the gap between domain-independent upper ontologies and the application-oriented domain ontologies has led to the proposal of a new kind of ontologies, viz. the top-domain ontologies, like the Simple Bio Upper Ontology [10], GFO-Bio [4], or BioTop [15]. Their purpose is to define the most general categories relevant for the entire biomedical domain, such as **Protein**, **Tissue**, **DNA**, or **BiologicalFunction** to be used as a common reference for more specialized domain ontologies.

¹ Corresponding Author: Stefan Schulz, Universitätsklinikum - Institut für Medizinische Biometrie und Medizinische Informatik, Stefan-Meier-Strasse 26, 79104 Freiburg, Germany, E-mail: stschulz@uni-freiburg.de

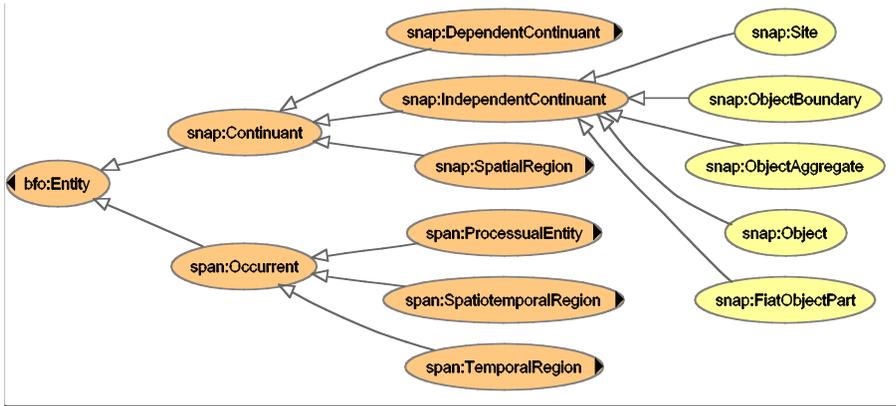


Figure 1. Excerpt from the BFO class tree

In the following we concentrate on the BioTop ontology² and its interface to BFO. The goal of BioTop is to provide classes and classificatory criteria to categorize the foundational kinds of biology, without any restriction to granularity, species, developmental stages or states of structural well- or ill-formedness [11]. In Bio Top's initial development, no definitive commitment existed towards any upper ontology, except for the distinction between continuants ("entities that endure through time while maintaining their identity and that are fully present at one single point in time", e.g., objects) and occurrents ("entities that are never fully present at any one given moment in time, but unfold themselves in successive phases, or temporal parts", e.g., processes), a principle on which most upper ontologies roughly agree. The primary focus at this point was primarily set on representing continuants from the area of interest. In the continued development the focus was broadened to include the representation of biological processes and BioTop was aligned with BFO. An analysis of the resulting implications of this alignment is at the center of the following deliberations.

1. Material and Methods

1.1. Upper Ontology Grounding

The choice of BFO was initiated by the following basic considerations:

- BFO's theoretical background had been extensively studied and scrutinized in logico-philosophical research on ontology for several years.
- An OWL-DL implementation of BFO was developed in a peer reviewed process and is freely available (with documentation) from its website³.

Figure 1 depicts a smaller subtree of BFO, highlighting the strict subdivision of the class `IndependentContinuant` with Table 1 summarizing the definitions of the specific classes which form the basis for the subsequent discussion. It is important to point out that BFO classes on the same level are supposed to be mutually disjoint (i.e., a given instance cannot instantiate sibling classes such as `Object` and `FiatObjectPart` at the same time). Further, from the definitions it becomes obvious that the notion of

² BioTop website: <http://purl.org/biotop>

³ BFO website: <http://www.ifomis.org/bfo>

“fiatness” [14] (e.g., the delimitation of a material entity without any physical discontinuity) is one of the most important differentiation criteria for the subclasses of IndependentContinuant.

BFO Class	Definition	Examples
Continuant	An entity that exists in full at any time in which it exists at all, persists through time while maintaining its identity and has no temporal parts.	
Independent-Continuant	A continuant that is a bearer of quality and realizable entities, in which other entities inhere and which itself cannot inhere in anything.	
Object	An independent continuant that is spatially extended, maximally self-connected and self-contained (the parts of a substance are not separated from each other by spatial gaps) and possesses an internal unity.	a cell, a body
Object-Aggregate	An independent continuant that is a mereological sum of separate object entities and possesses non-connected boundaries.	a heap of stones, a collection of bacteria
FiatObjectPart	An independent continuant that is part of an object but is not demarcated by any physical discontinuities.	the upper and lower lobes of the left lung, the dorsal and ventral surfaces of the body

Table 1. Textual definitions of the BFO classes used in this study

1.2. Ontology Modeling

The following principles have guided the development of BioTop: Firstly, full textual and logical definition of as many classes as possible to make the meaning of the various classes clear both to the human user as well as to automatic reasoning devices, and secondly, the avoidance of multiple is-a (inheritance) hierarchies because such hierarchies are intricately prone for modeling errors⁴.

BioTop is modeled in OWL-DL (Ontology Web Language - Description Logics) using the Protégé editor [5] together with the Pellet reasoner [13] for terminological reasoning and consistency checking. Language, editor, and reasoner have been established as standards in the Semantic Web community. During its development, the evolving ontology was repetitively assessed in terms of consistency (i.e., the classifier detected inconsistent classes which were then manually analyzed and corrected), and adequacy (i.e., domain experts checked the inferred hierarchy for correctness after each classification and corrected it, if necessary).

Once the ontology had reached the targeted coverage and both a consistent and adequate state, its topmost classes (e.g., EntireMolecularEntity,, HomogenousCollection, HeterogenousCollection, Organism) were mapped to BFO by importing its OWL-DL implementation and attaching BioTop classes as children to the respective BFO classes (e.g., MolecularFunction as subclass of BFO’s Function). In light of this integration, we encountered some phenomena which led us to critically appraise BFO’s differentiating criteria.

⁴ Given complete definitions, important taxonomic parents are later inferred by the terminological classifier anyway

2. Results

At a first glance, the inconsistencies found were mainly due to overlaps between the following subclasses of *IndependentContinuant*: *Object*, *FiatObjectPart*, *ObjectAggregate* (cf. Figure 1 and Table 1 for their respective BFO definition). As these classes are defined to be mutually disjoint, classification errors occurred wherever a *BioTop* class was found to fulfill the criteria to be subsumed by any two of these BFO classes (i.e., we did not assert multiple parents for a given *BioTop* class but rather let the reasoner infer such setting from the given set of logical restrictions on a class). An in-depth analysis yielded the following results:

2.1. *Object* vs. *ObjectAggregate*

BFO's alleged assumption that there is a clear ontological division between *Object* or *ObjectAggregate* is already challenged by the fact that any self-connected physical object can be also be described as a mereological sum of molecules, atoms, or elementary particles. A crucial point to stress here is self-connectedness: According to BFO, this property means that all the constituent parts are not separated from each other by spatial gaps. Albeit this is fully consistent with the restriction on *ObjectAggregate*, viz. that they are mereological sums of separate objects, the self-connectedness property requires a theory of connectedness in practice. For the ontology engineer there are two possibilities out of the dilemma:

Firstly, one could pursue an intuitive approach. On a macroscopic level spatial gaps can (possibly) be identified with the naked eye or by simple physical manipulation of the involved objects. But if shifting to a microscopic level and in case molecular entities should be represented, the situation might be quite different. Then, two cells could well appear disconnected under a light microscope but tiny interweaving molecular bridges could be revealed by a stronger electron microscope. As a consequence, modeling decisions are granularity-dependent.

Alternatively, one could resort to chemistry and the theory of the chemical bond in order to find a scientific criterion for connectedness or disconnectedness. Here continuous transitions [12] exist not only between the different kinds of bonds (i.e., ionic, covalent, metallic) but also in terms of strength of a bond [8]. For instance, large numbers of hydrogen bonds (which are in a strict sense not considered as molecule-forming chemical bonds) can form stable molecular complexes such as antigen/antibody aggregations. But on the other hand, the same kind of chemical bond can sporadically appear between separated biological membranes without any functional implications. Here, the case of liquids may be seen as borderline where the single molecules are not stationary but nevertheless cohesive due to connecting forces (e.g., water dipoles).

In *BioTop* we interpreted the connectedness requirement in a strict way only where no physical connecting forces can be assumed at all (e.g., in the instances of the classes *PortionOfGas* or *PopulationOfHumans*). This however entailed the necessity to create highly general *BioTop* classes (e.g., *Population*) as children of the union class *Object OR ObjectAggregate*, since there are populations of organisms in which the individuals are connected, such as in many aggregations of unicellular organisms (e.g., bacteria or yeast). But since *Object* and *ObjectAggregate* are disjoint, an instance of this union can therefore actually belong to either one of the classes but not to both at the same time.

2.2. Object vs. FiatObjectPart

With this distinction BFO postulates that everything can be classified according to the existence of a physical discontinuity (see again the definitions in Table 1), which is debatable, at least for biological entities. Beyond any doubt, many biologically relevant entities fulfill the strict criteria for **Object**, such as cells in the blood, molecules in a gas, or independent organisms. But many other classes cannot be unambiguously attributed to be either **Object** or **FiatObjectPart**. An emblematic example for this is a Siamese twin: Whereas the conjoined double-organism fulfills the criteria for **Object**, either twin is a **FiatObjectPart** because there is no clear-cut discontinuity between them. But after (surgical) separation, however, each twin becomes an independent **Object**. This discussion can be extended for many symbiotic organisms, to telophase cells in the mitotic cycle, or to atoms in some molecular associations. Finally, in the realm of gross anatomy, the conception of bodily organs kinds as **Object** would require to neglect their transitory structures (e.g., ducts, vessels, nerves) connecting them with their neighboring structures without any clear discontinuities.

In BioTop the emerging inconsistencies of classes subsumed both by **Object** and **FiatObjectPart** could only be fixed by introducing yet another union of BFO classes, *viz.* **Object OR FiatObjectPart**. Most important biological classes are subsumed by this union class (e.g., **Cell**, **Organism**, or **Atom**). But as with the union mentioned in the preceding section, the disjointness between **Object** and **FiatObjectPart** makes it impossible to declare e.g., the same given instance of a Siamese twin as a proper object or a fiat object part using the same ontology to just express different points of view.

3. Conclusion

Biomedical ontologies address different levels of granularity, reaching from populations down to single atoms. Neutrality towards granularity issues was a major architectural principle when creating BioTop, a top-domain ontology for the linking and interfacing of domain ontologies in the biomedical and biomolecular sciences.

The alignment of the BioTop classes with the Basic Formal Ontology (BFO) re-introduces some problems concerning granularity, since the BFO definitions that address aggregation and connection are obviously committed to a mesoscopic viewpoint and can therefore not be generalized to the whole range of granularity.

We here defend the opinion that upper ontologies should be neutral to granularity because addressing granularity issues in an upper ontology is like opening Pandora's Box: On the one hand this would presuppose a consensus on how to divide the granularity continuum, which is hard if not impossible to achieve. On the other hand, each level of granularity would then require its own set of axioms that could possibly clash with axioms from another granularity level. Also, to divide such cross-granular ontologies into separate ontologies tailored for granularity levels is likewise unfeasible since the inherently cross-granular nature of the biomedical domain. Furthermore, as the Siamese twin example demonstrates, the prominence of the notion of fiat boundaries complicates the classification of entities even at the same granularity level.

The conceptualization of connectedness is still an intricate philosophical problem [16]. Certainly, such problems may be addressed in some cases where a restricted domain description is committed to as well defined level of granularity. In such cases, the subdivisions of **IndependentContinuant** could indeed be quite useful. But for

BioTop, as a deliberately granularity independent ontology, both distinctions between Object vs. ObjectAggregate and Object vs. FiatObjectPart are rather inexpedient.

In accordance with several postings to the BFO discussion group⁵, where our report on the described problems was starting point of a lively (and ongoing) discussion, a tentative solution is emerging as follows:

1. For an extension of BFO in terms of a comprehensive domain top level (such as plugging BioTop to BFO) only the first two BFO levels are used.
2. In contrast, the mereotopological subclasses of IndependentContinuant should only be used once certain parameters (i.e., time, granularities) are fixed for a restricted subdomain.

Acknowledgments:

We thank all participants of the BFO Discussion Group for the lively exchange of ideas related to the discussed topics. Our work has been supported by the European Commission project BOOTStrep⁶ (FP6 - 028099).

References

- [1] A. Gangemi, N. Guarino, C. Masolo, and A. Oltramari. Sweetening Ontologies with DOLCE. In Proc. of the Internat.. Conference on Knowledge Engineering and Knowledge Management (EKAW), 2002.
- [2] P. Grenon, B. Smith, and L. Goldberg. Biodynamic Ontology: Applying BFO in the Biomedical Domain. In Proceedings of the Workshop on Medical Ontologies, pages 20-38, 2004.
- [3] B. Heller and H. Herre. Ontological Categories in GOL. *Axiomathes*, 14(1):57-76, 2004.
- [4] R. Hohndorf. GFO-Bio: A Biomedical Core Ontology, <http://onto.eva.mpg.de/gfo-bio.html>, last accessed: October, 15th, 2007.
- [5] H. Knublauch, O. Dameron, and M. Musen. Weaving the Biomedical Semantic Web with the Protege OWL Plugin. In U. Hahn, S. Schulz, and R. Cornet, editors, Proceedings of the 1st International Workshop on Formal Biomedical Knowledge Representation (KR-MED 2004), pages 39-47, 2004.
- [6] A. McCray, Y. Yang, and R. Patil. The UMLS Semantic Network. In Proc. of the 13th Annual Symposium on Computer Applications in Medical Care, pages 503-507, 1989.
- [7] L. Obrst, P. Cassidy, S. Ray, B. Smith, D. Soergel, M. West, and P. Yim. The 2006 Upper Ontology Summit Joint Communique. *Applied Ontology*, 1(2):203-211, 2006.
- [8] L. Pauling, editor. *The Nature of the Chemical Bond*. Cornell University Press, 1960.
- [9] A. Rector, J. Rogers, and P. Pole. The GALEN High Level Ontology. In Proc. of the 14th International Congress of the European Federation for Medical Informatics, pages 174—178, 1996.
- [10] A. Rector, R. Stephens, and J. Rogers. Simple Bio Upper Ontology, <http://www.cs.man.ac.uk/~rector/ontologies/sample-top-bio>, last accessed: October, 15th, 2007.
- [11] S. Schulz and U. Hahn. Towards the Ontological Foundations of Symbolic Biological Theories. *Artificial Intelligence in Medicine*, 39(3):237-250, 2007.
- [12] S. Schulz and I. Johansson. Continua in Biological Systems. *The Monist*, 90(4), 2007.
- [13] E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, and Y. Katz. Pellet: A Practical OWL-DL Reasoner. *Journal of Web Semantics*, 5(2), 2007.
- [14] B. Smith and A. Varzi. Fiat and Bona Fide Boundaries. *Philosophy and Phenomenological Research*, 60(2):401-420, 2000.
- [15] H. Stenzhorn, E. Beisswanger, and S. Schulz. Towards a Top-Domain Ontology for Linking Biomedical Ontologies. in Proc. of the 12th World Congress on Medical Informatics, pages 1225-1229, 2007.
- [16] A. Varzi. Mereological Commitments. *Dialectica*, 54:283-305, 2000.

⁵ BFO Discussion Group website: <http://www.ifomis.org/bfo/discussion>

⁶ BOOTStrep website: <http://www.bootstrep.eu>

Dichotomy – a forgotten ancient principle

György SURJÁN¹

National Institute for Strategic Health Research Budapest, Hungary

Abstract Dichotomy is an ancient principle of categorisation, where a class is divided into two jointly exhaustive and mutually disjoint categories. The principle as a general requirement was abandoned during the middle. The recent inquiry shows that studying this principle is still worthwhile and in some cases it can be used as a quality assessment tool. The paper presents algorithms that can transform any kind of categorial structures into dichotomy. The resulting representation sometimes can make apparent the problematic parts of the source. Problems often result from stating *Is_a* relations without differentiating criteria. A simple experiment of dichotomous transformation of the high level categories of the first chapter of ICD was carried out. The problem of "other" and "not elsewhere classified" categories is discussed. Conclusion: we should not strive to build dichotomous structures but sometimes a dichotomous transformation of an existing structure can be helpful to detect critical parts of a system of categories.

Keywords: Dichotomy, categorisation, ontology, quality assessment

Introduction

The principles of classification and formal ontologies first were discussed in detail by ancient philosophers, Plato and Aristotle. At their time one subject of debate was the dichotomy principle. In a strictly dichotomous classification each node has exactly two subclasses that are discriminated by the presence or absence of one definite property. The most famous ancient example of dichotomy is the so called Tree of Porphyry, an example of a strictly dichotomous top down hierarchy. [1] Aristotle himself was not very much in favour of dichotomy. He gave several arguments against dichotomy as a strict rule in one of his work 'De partibus animalium' [2]. In the 16th century Pierre de la Ramée (alias Petrus Ramus) provided a dichotomous system of logics (also called dialectics) [3].

Later on the dichotomy principle was abandoned and the author is not aware of any recent publication in the bio-medical informatics literature that supports this principle. It was recognised that strange structures can result from unreasonable enforcing of dichotomy. However, the aim of this paper is to show that it is still worthwhile to study and can play some role in modern bio-medical terminological systems, at least as a quality assurance tool. We definitely do not propose that current bio-medical terminologies or classifications should be dichotomous.

¹ Correspondence: György Surján ESKI Arany J u 6-8 Budapest Hungary 1051 e-mail: surjan.gyorgy@eski.hu

1. Definition of dichotomy and the dichotomous transformation

Dichotomy is splitting up a class into two jointly exhaustive and mutually exclusive subclasses. This means that if B and C are direct subclasses of A, then it is a dichotomous subdivision if

$$\forall x \text{ instance}(x, A) \Leftrightarrow \text{instance}(x, B) \vee \text{instance}(x, C) \text{ and}$$

$$\forall x \text{ instance}(x, A) \Rightarrow \neg(\text{instance}(x, B) \wedge \text{instance}(x, C))$$

Now we would like to show, that once we represent a hierarchical system of categories in a computer there is no need to decide in advance whether or not the system is dichotomous. This is because any hierarchical system that satisfies the criteria of a directed acyclic graph can be rewritten into a dichotomous form, although a number of new "auxiliary" categories have to be introduced.

Let say we have a tree structure, that has a node a with a series of subclasses a_1, \dots, a_n . If $n > 2$ the following procedure transforms this graph into a dichotomous structure:

Let us define an auxiliary category, a_{A1} that is defined as the union of a and $\neg a_1$. This will be a subcategory of a and a super-category of a_2 to a_n . In this new arrangement a has two direct subclasses, a_1 and a_{A1} . If a_1 has a differentiating criterion, then the presence or absence of that criterion makes the difference between the two subclasses of a . If a_{A1} has more than two subclasses then the same procedure should be repeated iteratively, until the whole structure is transformed to a dichotomous tree. This procedure can be applied to all nodes in a tree, so the whole graph can be transformed into a dichotomous structure. This procedure can be carried out algorithmically and a relatively simple computer program can do it within a predictable time. Therefore whenever there is a need for dichotomy, this can be established 'on the run'.

Let us call the procedure described above dichotomous transformation. Such a transformation preserves all categories and relations of the original structure but creates a number of new categories (what we called auxiliary categories) together with their relations.

Let we take granulocyte as an example. This is a kind of white blood cells that has three types, the so called basophil, eosinophil and neutrophil granulocytes according to staining property of the granules in the cytoplasm. This serves as a clear differentiating criterion for each type of granulocytes.

The transformation can be carried out in three different ways depending on which type we take first (Figure 1). This shows two important drawbacks of dichotomous structures. One is that it is quite accidental that a category appears at a higher or a lower level of the system. The other is that it is also accidental which auxiliary categories appear.

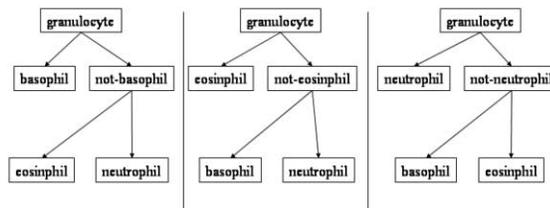


Figure 1

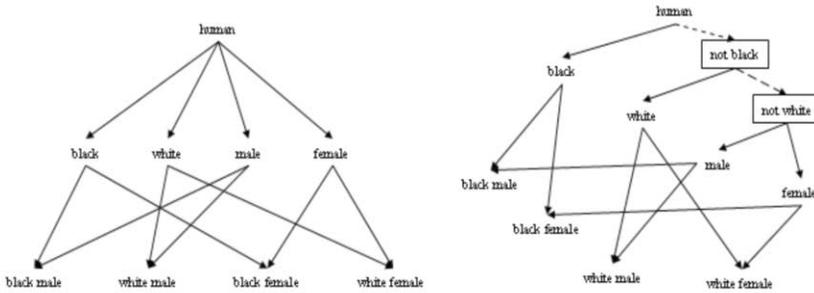


Figure 2

A blind, algorithmic dichotomous transformation of a system however might lead to false dichotomy. Let us take a system as an example that consists of the following categories: 'human', 'male', 'female', 'black' and 'white' (where black and white means human races), 'black male', 'black female', 'white male' and 'white female'. For sake of simplicity not other races are included.

The left side of Figure 2 illustrates a possible arrangement of these categories into a system, where the arrows represent 'Is-a' relations. A blind transformation of this system would lead to a rather fallacious system that is shown on the right side of Figure 2. The auxiliary categories are presented in boxes and the new relations by dotted lines. The fallacy in this arrangement is that 'male' and 'female' are kind of a 'not black' and 'not white' human; and this contradicts with the fact that e.g. 'black male' is a kind of 'black'. So 'black' and 'not black' have common subsumees, and the same holds for 'white' and 'not white'. This fallacy shows that something must be wrong in the original structure. Of course the mistake is that we used more than one differentiating criterion when we divided humans into 'male', 'female', 'black' and 'white' without any notice. We have to put either gender or race on the first level, so we get a consistent classification as shown in Figure 3. Both versions shown in this Figure are accidentally dichotomous, since we have two kinds of gender and also to kinds of race. A soon as we would include a third race, the structure would not be dichotomous any more, but still would be consistent, and each division would be jointly exhaustive and mutually exclusive. It is easy to see, that regardless the number of branches, a dichotomous transformation on such consistent structure always leads to a consistent dichotomy that is free from the fallacies that were shown above.

But if we put gender on the first level, we have to miss 'black' and 'white'; and if we put race first, we loose 'male' and 'female'. So the price of consistency is that we are not able to represent multiple inheritance any more.

2. Dichotomy and multiple inheritance

Multiple inheritance is an evergreen problem. In the age of paper based systems there was a technical constraint: there is no convenient way to represent multiple inheritance in large classifications (large here means that the whole system can not be drawn up on one single sheet). This technical limitation disappeared in the computer age, but later on it was recognised that manual building up poly-hierarchical structures is an error prone practice. Therefore it is often recommended today that *ontology engineers should*

build mono-hierarchies manually and the missing relations should be computed automatically based on formal definitions of the entities.

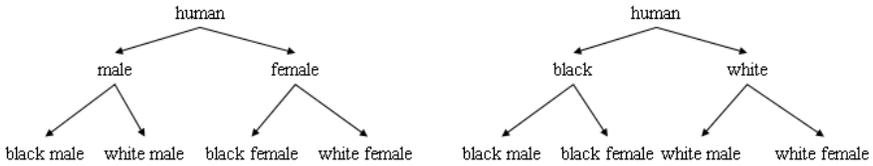


Figure 3

The only problem is that the nature of reality does not support this approach: I am male and white at the same time, and there are no obvious rules which of these relations should be asserted manually and which should be computed. But there are also no rules to choose from the options illustrated in Figure 3. More generally speaking: if we try to avoid poly-hierarchy by using strictly one differentiating criterion at each level, then there will be a number of equally justified solutions and we have to make an accidental decision by which some sensible categories (either male-female or black-white) will be lost.

The fallacy of the dichotomous transformation of the structure shown in Figure 2 makes clear that the assertion of *Is_a* relations without specifying the differentiating criterion is risky. That means that if we state that A subsumes B then we always have to tell, what makes B different from A. Doing so, it becomes clear in our example that e.g. 'black human' is different from 'human' due to the restriction on race, while 'female' is different from 'human' due to the restriction on gender. The conclusion is that instead of forcing a taxonomy structure, categories should be defined (e.g. a white male should be defined as a human who has male gender and white race). Building up the taxonomy is then a task that could be left to automated reasoning. Summing up this section, we propose to rephrase the above mentioned recommendation: *Assertion of an Is_a relation without specifying the differentiating criterion should be avoided.* The differentiating criterion, of course, consists of a *restricted property* and a *restriction value*. We can not claim that all divisions should be disjoint, because, as we saw, restrictions on different properties often lead to overlapping categories. We can not claim even that categories defined by restrictions on the same property should be disjoint. In case of gender, 'male' and 'female' are apparently disjoint; but there are many other properties for which an entity can have more than one value. E.g. scientific papers can be classified according to their authors. Since several papers are co-authored, a restriction on the 'has_author' property will not lead to disjoint categories.

If the proposed recommendation is followed, the dichotomous transformation is always possible without resulting fallacious structures. Instead of the described blind method, we can transform such systems into a decision tree forming a YES/NO question from each property (E.g. "Is this human a female?" or "Is this paper authored by Surján?"). Each possible value of each restricted property should be transformed to such a question. The series of all questions forms a dichotomous tree. Since categorical structures rarely contain all possible combinations, some of the questions could be skipped depending on the answer to previous questions.

3. Dichotomy principle as a quality assessment tool, an experiment with ICD

The above described method of dichotomous transformation can be used as a quality assessment tool, since if a system contains ill-defined categories, the transformation fails or result in inconsistency. We tested this on the upper level categories of the first chapter of ICD 10 (Certain infectious and parasitic diseases). Some parts of the chapter that are not infections were intentionally omitted. The result of this manual transformation is shown in Figure 4. If one goes through the questions, sooner or later one arrives at a code range or section (e.g. A80-A89) that can be considered as a root of another tree recursively; by moving along the branches of such a sub-tree one can finally arrive at a definite three digit category of ICD. (Again in turn the three digit code can be seen as the root of a third level tree that leads to four-digit codes). Wherever there is a (!) mark in the figure, there is some serious problem with the approach.

At the first line the problem is that the title of the 1st chapter in ICD is not "Infectious diseases" but "*Certain* infectious and parasitic diseases", revealing the fact, that there are infections in other chapters also. See also A20-A28: *certain* zoonotic diseases. In case of spirochete and Chlamydia infections the word 'other' refers to the fact that again some of these infections belong to other parts of the system.

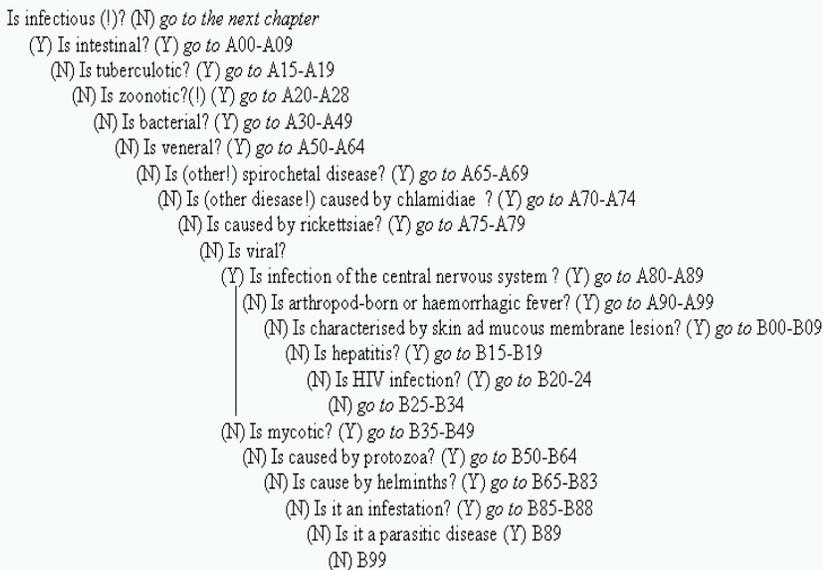


Figure 4

There is an implicit rule here that should be followed. Once we arrive to a section and go through the corresponding sub-tree, it may happen that none of the questions in the sub-tree is answered positively. Then we ought to return to the first level, and continue. But in case of the sexually transmitted diseases, before we could return to the upper level, we find A48 'Other bacterial diseases not elsewhere classified'. This means, that if we got a YES to the question 'Is bacterial' and arrive to A48, this branch should be first skipped, all other possibilities have to be investigated and only if all other

questions get NO, can we return and use A48. Similar problem would rise in many of the sections of ICD. Cimino [4] – confirmed by Rector [5] – suggest to abandon the “not elsewhere classified” categories since they have no definite semantics: not being classified elsewhere is not an intrinsic property of an entity, rather a feature of a given classification. What we can add to this based on the above described experience with dichotomous transformation, that these categories can make any decision algorithm rather difficult.

4. Conclusions

The dichotomy principle was an idea of some ancient philosophers, and there was no reason to maintain it over centuries. In the computer era there is no reason to follow the dichotomy principle when designing a terminological system, since it would be not convenient in many cases. Whenever it is needed it is possible to convert any tree structure into a dichotomous tree. Even lattice structures – if all entities have well defined differentiating criteria – can be transformed into dichotomous decision trees. However the dichotomy principle is something that is still worthwhile to keep in mind, as something that can be useful in testing and improving quality of classifications and ontologies. Quality management is a hot issue in this field even today [6] and in respect of newly developed systems [7], [8]. This is what gives still relevance to the dichotomy principle even today. This study of dichotomy has shown that stating *Is_a* relations between categories without defining the restricted property of the subsumed category is something that should be avoided, because mixing up restrictions on different properties might lead to errors.

5. Acknowledgement

The author expresses his special thanks to Arie Hasman, Gergely Héja and members of Health Ontology Forum (www.eski.hu/eof) for their valuable comments and remarks on this paper.

References

- [1] Ross, G (n.d.) Porphyry, Introduction to Aristotle's categories Retrieved June 9, 2007 from <http://www.philosophy.leeds.ac.uk/GMR/hmp/texts/ancient/porphyry/isagoge.html>.
- [2] Ross, D. . *Aristotle* London: Methuen & Co Ltd. 1977
- [3] <http://etext.virginia.edu/cgi-local/DHI/dhi.cgi?id=dv4-06>. Last accessed: February 25th, 2008.
- [4] Cimino J J Desiderata for controlled medical vocabularies in the twenty-first century *Met Inf Med* 1998;37:394-403
- [5] Rector AL. Clinical terminology: why is it so hard? *Methods Inf Med.* 1999;38(4–5):239–52.
- [6] Ceusters W, Smith B, Kumar A, Dhaen C. Mistakes in medical ontologies: where do they come from and how can they be detected? *Stud Health Technol Inform.* 2004;102:145-63.
- [7] Chiang MF, Hwang JC, Yu AC, Casper DS, Cimino JJ, Starren JB. Reliability of SNOMED-CT coding by three physicians using two terminology browsers. *AMIA Annu Symp Proc.* 2006;:131-5.
- [8] Schulz S, Suntisrivaraporn B, Baader F. SNOMED CT's problem list: ontologists' and logicians' therapy suggestions. *Medinfo.* 2007;12(Pt 1):802-6.

12. Doctoral Consortia Papers

This page intentionally left blank

Attention and usability issues in mobile health information systems at point-of-care

Ole Andreas ALSOS

The Norwegian EHR Research Centre & Department of Computer and Information Science, Norwegian University of Science and Technology

Abstract. In point-of-care situations, use of mobile devices may demand much of the physician's attention and may disturb the communication with the patient. By using minimal attention user interface design and context awareness, attention theft from mobile devices at point-of-care can be reduced.

Keywords: Usability, Health Information Systems, Human-Computer Interaction, User-computer interface.

1. Introduction

1.1. Problem description

While the use of mobile devices is becoming reality in many clinical settings, little is known about the effect of their introduction on the physician-patient communication in point-of-care situations. A mobile device may act as a support to the clinician when he communicates with the patient, but may also draw the attention of the physician away from the patient – a situation that hereafter will be described as 'attention theft'.

Occurrences of attention theft have been reported from observations of computing device usage in clinical environments, both for laboratory [1] and real-life settings [2].

Relevant contributions for attention-easy mobile user interfaces emerge from different areas within the field of human computer interaction. *Minimal attention user interfaces* [3], *Context aware applications* [4] and *Attentive User Interfaces* [5] might provide a relevant reduction in cognitive workload.

1.2. Research questions

Our main goal is to develop usability guidelines for attention-easy and non-disruptive mobile information systems at point-of-care. Towards that goal we compare demands for a mobile device and a medical paper chart used in point-of-care situations.

2. Methodology

Lab experiments (N=14) in a full-scale model of a section of a hospital ward were used in the study. During a simulated patient visit physicians made changes in the prescribed medication by using four different information devices; one paper based medical chart and three versions of a PDA-based medication system. Three different input methods were used for the PDA (stylus, finger and buttons).

3. Results

3.1 Preliminary results

As a first step towards creating a framework for measuring attention theft, we analyzed gaze direction times from video data from a previously conducted usability evaluation [3]. In this test physicians and patients playing out a clinical scenario with and without using a PDA for selecting x-ray images for display on a bed-side mounted screen. The results showed that PDA usage increased the number of focus shifts by 87% and increased the number of slowed-speech events by 108%.

The posttest debrief after the lab study indicate that a majority of physicians are positive towards mobile systems at point-of-care, but worried about attention theft and the ICT system's effect on the communication with the patient.

3.2 Current status

Laboratory experiments are performed and video data is currently being analyzed.

4. Discussion

4.1. Discussion of the results

Because focus shifts incur cognitive costs [6] and slowed speech indicates resource depletion caused by increased cognitive demands [7], our results indicate that physicians experience attention theft from mobile point-of-care system. Careful design of the user interface might minimize the attention required for these systems.

4.2 Plan for future work

We will develop a framework for measuring attention theft, analyze observational data, and develop guidelines for attention-easy user interfaces in point-of-care situations.

5. Contribution to the field of Biomedical informatics

With the rapidly increasing computer support for care-critical functions in hospitals, mobile devices with well-tuned user interfaces are needed to avoid unnecessary attention theft. From a general point of view the ability to design attention-easy interfaces for mobile work will contribute to a better user experience.

References

- [1] Alsos, O.A. and D. Svanaes. Interaction techniques for using handhelds and PCs together in a clinical setting. in NordiCHI. 2006.
- [2] Bardram, J., The trouble with login: on usability and computer security in ubiquitous computing. Personal and Ubiquitous Computing, 2005. 9(6): p. 357--367.
- [3] Pascoe, J., N. Ryan, and D. Morse, Using while moving: HCI issues in fieldwork environments. ACM Transactions on Computer-Human Interaction, 2000. 7(3): p. 417--437.
- [4] Kristoffersen, S. and F. Ljungberg. "Making place" to make IT work: empirical explorations of HCI for mobile CSCW. in GROUP '99: Proceedings of the international ACM SIGGROUP conference on Supporting group work. 1999. New York, NY, USA: ACM Press.
- [5] Vertegaal, R., Designing attentive interfaces. 2002, New Orleans, Louisiana: ACM Press. 23-30.
- [6] Welford, A.T., The "psychological refractory period" and the timing of high-speed performance - A review and a theory. 1952. 43(2).
- [7] Lund, N., Attention and pattern recognition, ed. C. Flanagan and P. Banyard. 2001, Hove: Routledge.

Reconstructing Clinical Events by Interpreting NICU Monitoring Data

Feng GAO

Department of Computing Science, University of Aberdeen¹

Abstract. Several studies are reported in literature that use electronically recorded data for aiding decision support/reporting in the Neonatal Intensive Care Unit (NICU). In this PhD research I investigate techniques to automatically reconstruct clinical events that are not recorded directly but could be reconstructed from continuous physiological data and other discrete data.

Keywords: Bioinformatics, pattern recognition, Decision support, Data analysis-extraction tools, Classification

1. Introduction

1.1. Problem description

Computer systems are used within many ICUs to collect and record patient medical data [1] with a view to improving patient care [2-3]. In particular, the BabyTalk project [4] aims at automatically analysing Neonatal ICU data to generate textual reports for decision support and reporting. My PhD research is part of the BabyTalk project.

1.2. Research questions

I aim to explore the reconstruction of important clinical events (bradycardia, intubation, etc - which are not normally recorded) by interpreting continuous physiological monitoring data and other discrete data. This is necessary if we are to generate a more complete picture of what happened at what time to the patient. I will need to answer questions like:

1. What are the important clinical events? Does the set of clinical events differ according to the purpose of the reconstruction (decision support, reporting, etc)?
2. How difficult it is to reconstruct the important events? Is the existing monitoring data sufficient to reconstruct them?

2. Methodology

My research will involve knowledge acquisition from medical experts to identify the

¹ Adress for correspondence: fgao@abdn.ac.uk

important events from the BabyTalk ontology [4]. I also need to investigate how experts identify these events and to establish the difficulty of automatic reconstruction.

Our exploratory studies of medical events with experts revealed that these events do not always manifest themselves as observable patterns in the physiological monitoring data. Therefore, our methodology involves developing techniques that analyse these and other available discrete data to build medical contexts that help in reconstructing required events.

3. Preliminary Results and Current Status

So far, I have established a small set of medical events with different scores for importance and difficulty. Bradycardia is included as an event with low importance and low difficulty score. A system using sliding windows based data abstraction and decision tree induction has been implemented to generate rules for reconstructing bradycardia. Currently I am testing segmentation as a data abstraction method and using segmentation with machine learning to generate the knowledge base.

4. Discussion

The performance [5] of a learned decision tree is as good as the current method used in BabyTalk prototype. It is better at the boundary estimation. However sliding window based data abstraction is not very comprehensible to human being.

In the future, the event set and their importance/difficulty score need to be refined. Interval based data abstraction and multi-dimensional data analysis will be applied to reconstruct more complicate events. Different machine learning methods will be tested in generalizing the medical knowledge base for event reconstruction.

5. Contribution to the field of Biomedical informatics

Prioritising medical events that need to be reconstructed for reporting or for decision support. Developing and evaluating computational techniques that exploit expert medical knowledge for reconstructing the prioritised events.

References

- [1] Green CA, Gilhooly KJ, Logie RH, Ross DG. Human factors and computerisation in Intensive Care Units: a review. *Int J Clin Monit Comput* 1991; 8: 167-178.
- [2] Jim Hunter, Neil McIntosh: Knowledge-based event detection in complex time series data. AIMDM'99, Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making, Springer Verlag LNCS 1620, pp 271-280, 1999.
- [3] Christine L. Tsien, Isaac S. Kohane, Neil McIntosh: Multiple signal integration by decision tree induction to detect artefacts in the neonatal intensive care unit. *Artificial Intelligence in Medicine* 19 (2000) 189-202.
- [4] François Portet, Ehud Reiter, Jim Hunter and Somayajulu Sripada. Automatic Generation of Textual Summaries from Neonatal Intensive Care Data. 11th Conference on Artificial Intelligence in Medicine (AIME 07), pp. 227-236, 2007
- [5] F. Portet, F. Gao, J. Hunter, Y. Sripada. Evaluation of On-Line Bradycardia Boundary Detectors from Neonatal Clinical Data. 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society

Evaluating the Impact of CPOE Systems on Medical Workflow: a Mixed Method Study

Zahra NIAZKHANI

Institute of Health Policy and Management (iBMG), Erasmus University Medical Center, Rotterdam, the Netherlands

Abstract. For a successful design and deployment of computerized provider order entry (CPOE) systems which support medical workflow, thorough understanding of workflow and its changes after the CPOE implementation are crucial. For this purpose, this PhD study incorporates insights from multiple disciplines and uses mixed research methods to study a CPOE implementation case at a Dutch hospital. The study evaluates the workflow of different user groups in different phases of the implementation. The results have shown that the system impacted workflow both positively and negatively. Although providers exploited its supportive features in workflow, they also needed to devise a number of workarounds to overcome its simultaneously introduced barriers in workflow.

Keywords: medical order entry systems, medical work, CPOE, workflow

1. Introduction

The potential of CPOE systems to change provider's workflow and its consequence for patient safety has brought the issue of workflow to the forefront of CPOE implementations [1]. For a successful deployment of these systems, thorough understanding of workflow and its changes after CPOE implementation are crucial. In this PhD study, I address three main questions: Which aspects of medical workflow are most affected by CPOE implementation and how? Which factors do influence these effects? And how do providers integrate the system into their work practice?

2. Methodology

This study uses multiple research approaches to study a CPOE implementation case at a Dutch hospital. Insights from the social sciences, the field of Computer Supported Cooperative Work, the information sciences, and medical informatics served as theoretical background to study workflow with a CPOE system. The empirical data included both quantitative and qualitative data. To evaluate user attitudes towards workflow, three questionnaire surveys were conducted: two weeks before and approximately five months and three years after the implementation. The qualitative approach used 23 key informant interviews with different professional groups supplemented with process mapping in the medication ordering and administration process and analysis of handwritten records and system printouts used in this process.

3. Results

A quantitative, before-after comparison showed that the view of providers on the system's overall impact on work depends to a large extent on the change of their paper-based work structure [2]. With regard to support of workflow, no difference was found between the paper-based and the computerized systems. Moreover, a literature review of CPOE evaluation studies showed that modeling principles of CPOE systems generally make use of the formal, predefined division of tasks and preconceived relationships between clinical tasks and also between healthcare professionals [3]. In the workflow model underlying these systems, the collective and collaborative nature of workflow is greatly neglected, although in practice workflow is largely negotiated and co-constructed among members of a care team. This finding was confirmed by a qualitative case-study of workflow between different professional groups in which several problematic workflow integration issues were identified [4]. To understand how providers overcome workflow barriers after the implementation, an in-depth analysis of workflow in the medication ordering and administration process was performed [5]. This led to uncovering of a number of workarounds devised to bypass the barriers in workflow.

4. Discussion

In this study, the CPOE system affected workflow both positively (e.g., legible and complete orders) and negatively (e.g., problematic collaboration between providers). Therefore, although providers exploited the system's supportive features in workflow, they also needed to devise a number of workarounds to overcome its simultaneously introduced barriers in workflow. In practice, despite problems introduced by these systems in communication and collaboration [6], they may continue to operate only because providers devise workarounds to bypass its barriers. As these workarounds are not registered by the system, they may be simply disregarded in system design or redesign. Therefore, only studying the CPOE system closely in real practice can help to better recognize and understand its actual behaviour and true impact on workflow. The insights gained in this study can help to bridge gaps between CPOE design principles and the social organization of workflow.

References

- [1] Campbell EM, Sittig DF, Ash JS, Guappone KP, Dykstra RH. Types of unintended consequences related to computerized provider order entry. *J Am Med Inform Assoc.* 2006 Sep-Oct;13(5):547-56.
- [2] Niazkhani Z, Van der Sijs H, Pirnejad H, Redekop W, Aarts J. Same system, different outcomes: comparing the transitions from two paper-based systems to the same CPOE system. *Submitted for review.*
- [3] Niazkhani Z, Pirnejad H, Berg M, Aarts J. The impact of CPOE systems on medical workflow: a literature review. *To be submitted.*
- [4] Niazkhani Z, Pirnejad H, De Bont A, Aarts J. Evaluating inter-professional work support by a computerized physician order entry system. *Accepted paper for oral presentation at MIE2008.* Goteborg, Sweden May 25-28, 2008.
- [5] Niazkhani Z, Pirnejad H, Aarts J. Opening the black box: workarounds to integrate a CPOE system into practice. *Paper submitted for review to AMIA 2008 Annual Symposium*
- [6] Pirnejad H, Niazkhani Z, Berg M, Bal R. Intra-organizational communication in healthcare: considerations for standardization and ICT application. *Accepted for publication in Methods Inf Med.*

Author Index

Aarts, J.	v, 321	Bjorkdahl, A.	77
Adlassnig, K.-P.	121	Blobel, B.G.M.E.	359, 629, 673, 697, 709, 735
Agorastos, T.	241	Blomstrand, C.	77
Åhlfeldt, H.	401	Bodemer, C.	51
Ahmadian, L.	127	Bødker, K.	389
Airey, C.	496	Boeker, M.	863
Allaert, F.-A.	667	Borch, N.	83
Alsos, O.A.	541, 877	Borycki, E.M.	505, 567
Ammenwerth, E.	461, 771	Botsis, T.	365
Andersen, R.	57	Böttger, S.	63
Andersen, S.K.	v	Bouaud, J.	139
Andersen, U.	353	Bourdon-Lanoy, E.	51
Anguita, A.	3	Bousquet, C.	145, 803, 857
Årsand, E.	113	Boussadi, A.	145
Austin, T.	685	Boyer, C.	407
Avillach, P.	205	Braithwaite, J.	15
Ayme, S.	51	Brattheim, B.	371
Azaouagh, A.	479	Bratvold, A.	83
Bakhshi-Raiez, F.	779	Brender, J.	611
Bakker, P.J.M.	303, 327	Brezinka, V.	71
Balas, A.	797	Briançon, S.	605
Bamidis, P.D.	517, 653	Bricon-Souf, N.	413
Bari, M.	839	Broeren, J.	77
Bar-Or, D.	253	Brown, S.H.	797
Bar-Or, R.	253	Bryhni, T.	57
Bassøe, C.-F.	133	Bryhni, H.	57
Bath, P.A.	297	Buchtela, D.	151, 377
Baujatz, G.	51	Burgun, A.	455
Bayat, S.	605	Bürkle, T.	229, 279
Beale, T.	401	Burkow, T.M.	83
Beck, P.	473	Busch, C.	673
Beisiegel, T.	827	Buyl, R.	383
Beisig, A.	279	Cadonna, B.	473
Beisswanger, E.	9	Callen, J.	15
Bellika, J.G.	271, 365	Carrajo, L.	395
Belsis, P.	661	Cavallini, A.	573
Ben Said, M.	51	Chen, R.	401
Bergstrøm, R.	617	Claesson, L.	77
Bernad, E.	839	Clarke, M.	181, 717
Bernstein, K.	353	Coatrieux, G.	667
Beuscart, R.	413, 647	Cohen, G.	21
Beuscart-Zephir, M.-C.	413	Cohen, R.	339
Beuteführ, H.	63		

Colaert, D.	641	Framiñán Torres, J.M.	511
Colombet, I.	145	Franco, A.	39
Cornet, R.	127, 779, 785	Frimat, L.	605
Courtney, K.L.	555	Froehling, D.	797
Couto, E.	395	Fursse, J.	181
Crespo, J.	3	Füssle, R.	63
Cruchet, S.	407	Gabin, J.	617
Cuggia, M.	605, 815	Gall, W.	425
Dahamna, B.	33, 205, 235, 467	Gamberger, D.	851
Dahlström, Ö.	157	Ganascia, J.-G.	547
Dameron, O.	9	Ganslmayer, M.	279
Darmoni, S.J.	33, 205, 235, 467, 845	Gao, F.	879
Das, A.	541	Garcia-Remesal, M.	163
Daskalakis, S.	285	Garde, S.	401, 437
de Bont, A.	321	Gaudinat, A.	407
de Keizer, N.F.	127, 779	Geissbuhler, A.	21, 523
de la Calle, G.	163	Georgiou, A.	291, 345
Degardin-Capon, N.	413	Gerard-Blanluet, M.	51
Degoulet, P.	145	Gibert, K.	95
Deken, V.	647	Gindl, S.	187
Del Río Ortega, J.A.	511	Giorgi, I.	579
Deleger, L.	89	Glück, E.	617
Delopoulos, A.	241	Gomes, R.	765
Denecke, K.	791	González Sánchez, M.J.	511
Derville, A.	845	González, F.J.	395
Dhakal, S.B.	535	Goodwin, L.K.	555
Diou, C.	241	Goud, R.	193
Dorda, W.	425	Goude, D.	77
Dostalova, T.	529	Grabar, N.	199, 809
Dufour, E.	51	Granlien, M.F.	389, 419
Dufts Schmid, G.	425	Grossmann, W.	425
Duguay, C.	39	Groth, M.	333
Duhamel, A.	647	Gudmundsen, J.	419
Ebrahiminia, V.	223, 339	Gusew, N.	623
Eklund, P.	759	Gustafsson, M.	175
Eldevik, O.P.	535	Gütl, C.	473
Elkin, P.L.	797	Hägglund, M.	597
Ellini, A.	223	Hahn, U.	9
Elvidge, K.	169	Hailes, S.	685
Estacio-Moreno, A.	803	Hamon, T.	809
Falcoff, H.	223	Handels, H.	333
Falelakis, M.	241	Hanzlicek, P.	529
Falkman, G.	175	Happe, A.	815
Fassa, M.	667	Hartmann, B.	63
Faxvaag, A.	371	Hartvigsen, G.	113, 271, 365, 535
Fest, T.	455	Hasman, A.	193
Fieschi, M.	205	Hass, U.	157
Findlay, G.	181	Hasvold, P.	535
Finozzi, E.	579	Haux, R.	449, 623

Heimly, V.	431	Koufi, V.	679
Héja, G.	821	Koutkias, V.	241
Hejlesen, O.	365	Krogstad, T.	83
Hellrung, N.	623	Kumar, A.	857
Herbst, M.	673	Kushniruk, A.W.	567
Hersh, W.	523	Kuwata, S.	567
Hertzum, M.	419	Laila, M.	39
Hildebrand, C.	629	Lamy, J.-B.	223
Hilwig, H.	247	Landais, P.	51
Hippman, R.	529	Lang, M.	229
Hitoglou-Antoniadou, M.	517	Lapão, L.V.	765
Holban, S.	839	Laumann, S.	229
Hörbst, A.	771	Lauteslager, A.	303
Horhat, R.F.	561	Le Beux, P.	605, 815
Hovsto, A.	709	Le Merrer, M.	51
Hsu, W.-C.	297	Le Mignot, L.	51
Imbriani, M.	579	Le Toumelin, P.	339
Ingenerf, J.	827	Lea, N.	685
Itälä, T.	723	Lecroq, T.	235
Jais, J.-P.	51	Lee, V.	9
James, A.G.	833	Leitner, H.	473
Jaspers, M.W.M.	193, 303, 309, 327	Lekka, I.	241
Jalent, M.-C.	199	Lemieux-Charles, L.	505
Jilani, I.	199	Leonardi, G.	573
Jones, R.	181	Letord, C.	235
Jontell, M.	175	Levy, P.P.	547
Joubert, M.	205, 235	Lewalle, P.	635
Jovic, A.	851	Liebow, M.	797
Juhola, M.	211	Lind, L.	101
Kaiser, K.	187	Lindgren, H.	315
Kajbjer, K.	729	Livartowski, A.	809
Kalpathy-Cramer, J.	523	Lopez Alonso, V.	45
Kalra, D.	635, 685	Lopez, D.M.	735
Kanter, A.S.	27	Lopez-Campos, G.	45
Karlsson, D.	401, 753	Lovis, C.	641
Kasparova, E.I.	265	Lövström, R.	753
Kerdelhué, G.	33	Lundgren-Nilsson, Å.	77
Kessler, M.	605	Luneski, A.	517
Khajouei, R.	309	Lungeanu, D.	839
Khemka, S.	181	Mabeck, H.	443
Kim, J.-J.	9	Maglaveras, N.	241, 653
Kissoon, N.	247	Mair, R.	771
Klasen, J.	63	Malamateniou, F.	747
Klein, G.O.	v, 401	Mans, R.	573
Knaup, P.	437	Mantas, J.	285
Koch, S.	597	Maojo, V.	163
Kohl, C.D.	437	Marschollek, M.	449
Kokkinakis, D.	217	Martín, L.	3
Koster, P.	484	Martin-Sanchez, F.	45

Masarie, F.E.	27	Pereira, S.	235, 845
Massari, P.	845	Petković, M.	484
Mazzoleni, M.C.	v, 579	Pettersson, M.	753
McNair, P.	611	Peute, L.W.P.	303, 327
Melcón, R.	395	Pharow, P.	359, 629, 673, 697, 709
Melleby, O.-F.	617	Philippe, C.	547
Meneton, P.	199	Pietrasz, S.	547
Merabti, T.	235	Pirnejad, H.	321
Messiaen, C.	51	Plischke, M.	449
Mihalas, G.I.	561	Poland, G.	797
Miils, E.	496	Popescu, M.	555
Mikalsen, M.	107	Prcela, M.	851
Miksch, S.	187	Prokosch, H.-U.	229, 279
Mitkas, P.	241	Psutka, J.	529
Mizushima, H.	585	Pugliese, F.	579
Montagner, J.	667	Quaglini, S.	573
Morse, W.	291	Quantin, C.	667
Mortensen, W.	113	Rakovac, I.	473
Müller, H.	523	Rappelsberger, A.	121
Müller, T.H.	741	Rath, A.	51
Müllerleile, K.	333	Ray, S.	291
Mykkänen, J.	723	Rebholz-Schuhmann, D.	9
Naeymi-Rad, F.	27	Rector, A.	635
Nagy, M.	529	Relan, J.	333
Neamtu, M.	561	Rialle, V.	39
Neubauer, T.	691	Riaño, D.	95
Niazkhani, Z.	321, 881	Richard, J.-B.	51
Nicolas, L.	39	Riedl, B.	691
Niczko, E.J.	63	Risberg, M.J.	83
Nielsen, A.L.	247	Rizand, P.	809
Nieto Cervera, J.	511	Rob, M.I.	345
Noaghiu, R.	839	Roberts, J.M.	591
Nyssen, M.	383	Rodrigues, J.M.	635, 857
Nyström, M.	401, 753	Rognoni, C.	579
Nytrø, Ø.	703	Röhrig, R.	63
Ohashi, W.	585	Rosenbloom, S.T.	797
Olsen, B.I.	535	Rossille, D.	455
Olsen, O.-A.	113	Røstad, L.	703
Østengen, G.	83	Roux, C.	667
Pagani, M.	579	Ruland, C.M.	57
Pangault-Lorho, C.	455	Ryan, A.	759
Pantziou, G.	661	Rydmark, M.	77
Panzarasa, S.	573	Sabatier, B.	145
Papakonstantinou, D.	747	Saboor, S.	461
Parra Calderón, C.L.	511	Safran, C.	27
Pearl, A.	253	Salleron, J.	647
Peek, N.	193	Salomon, R.	51
Peleška, J.	151, 377	Samuelsson, H.	77
Penas, Á.	395	Säring, D.	333

Savastano, M.	709	Trombert, B.	857
Sax, H.	21	Trusko, B.	797
Scandurra, I.	597	Tsiknakis, M.	3
Schabetsberger, T.	771	Turlin, B.	815
Schmidt, R.	259	Turner, P.	490
Schonenberg, H.	573	Ugon, A.	547
Schulz, S.	v, 9, 863	Ukkola, J.	723
Seim, A.R.	371	Ustun, B.	635
Séroussi, B.	139	Valls, A.	95
Seydlova, M.	529	van der Aalst, W.	573
Shifrin, M.A.	265	van Deursen, T.	484
Skogh, T.	157	van Klei, W.A.	127
Skourlas, C.	661	van Vuurden, K.	271
Smidl, L.	529	Varga, P.	821
Sögner, P.	771	Varmedal, R.	113
Soualmia, L.F.	467	Varoutas, P.-C.	809
Spackman, K.A.	833	Vartzopoulos, D.	653
Spat, S.	473	Vassilacopoulos, G.	679, 747
Spithoven, R.	327	Vassiss, D.	661
Splendiani, A.	9	Venot, A.	223, 339
Spyrou, S.	653	Verloes, A.	51
Stark, G.	473	Vesely, A.	151
Stausberg, J.	479	Vigil Martín, E.	511
Stav, E.	107	Virtanen, A.	723
Stenzhorn, H.	863	Virtanen, M.	635
Stewart, S.	555	Vognild, L.K.	83
Stroetmann, V.N.	635, 641	Vorobieva, O.	259
Sulman, C.	647	Wahner-Roedler, D.	797
Sundvall, E.	753	Walderhaug, S.	107
Sunnerhagen, K.S.	77	Wang, A.Y.	27
Surján, G.	635, 821, 869	Watanabe, H.	567
Svanæs, D.	541	Westbrook, J.I.	15, 291, 345
Symeonidis, A.	241	Wihlborg, J.	753
Tanaka, H.	585	Willkomm, M.	623
Teelucksingh, S.	247	Wolf, K.-H.	449
Tegtbur, U.	449	Wong, M.C.	490
Temesgen, Z.	797	Wozak, F.	771
the CEMARA Task Force	51	Wrba, T.	425
Thirion, B.	33, 845	Yee, K.C.	490, 496
Thyberg, I.	157	Zaharie, D.	839
Timmins, W.	291	Zanstra, P.	635
Timpka, T.	157	Zucker, J.-D.	223
Torgersson, O.	175	Zvárová, J.	151, 377, 529
Toussaint, Y.	803	Zvolský, M.	151, 377
Toussi, M.	339	Zweigenbaum, P.	89
Trmal, J.	529		

This page intentionally left blank