HEALTHINF 2008

International Conference on
Health Informatics

Proceedings

Volume 2

Funchal, Madeira - Portugal · 28 - 31 January, 2008

TECHNICAL CO-SPONSORSHIP BY

IEEE EMB

MEMBER Workflow Management Coalition WfMC

CO-ORGANIZED BY

INSTICC

UNIVERSIDADE da MADEIRA

IN COOPERATION WITH

aaai

INSTICC PRESS

HEALTHINF 2008

Proceedings of the
First International Conference
on Health Informatics

Volume 2

Funchal, Madeira - Portugal

January 28 – 31, 2008

Co-organized by
INSTICC – Institute for Systems and Technologies of Information,
Control and Communication
and
UMA – University of Madeira

Technical Co-sponsorship by
IEEE EMB – Engineering in Medicine and Biology Society
and
WfMC – Workflow Management Coalition

In Cooperation with
AAAI – Association for the Advancement of Artificial Intelligence

Copyright © 2008 INSTICC – Institute for Systems and Technologies of Information, Control and Communication
All rights reserved

Edited by Luís Azevedo and Ana Rita Londral

Printed in Portugal

ISBN: 978-989-8111-16-6

Depósito Legal: 268554/07

http://www.healthinf.org

secretariat@healthinf.org

HEALTHINF is part of BIOSTEC – International Joint Conference on Biomedical Engineering Systems and Technologies

BRIEF CONTENTS

INVITED SPEAKERS

Sérgio Cerutti

Polytechnic University of Milan

Italy

Kevin Warwick

University of Reading

U.K.

Fernando Henrique Lopes da Silva

University of Amsterdam

The Netherlands

Vipul Kashyap

Partners HealthCare System, Clinical Informatics R&D

U.S.A.

David Hall

Research Triangle Institute in North Carolina

U.S.A.

Albert Cook

University of Alberta, Faculty of Rehabilitation Medicine

Canada

SPECIAL SESSION CHAIRS

SPECIAL SESSION ON KNOWLEDGE DISCOVERY AND DECISION SUPPORT SYSTEMS IN HEALTH INFORMATION SYSTEMS

Domenico Conforti, Università della Calabria, Italy

Manolis Tsiknakis, Foundation for Research and Technology - Hellas, Institute of Computer Science, Greece

Joerg Habetha, Philips Technologie GmbH, Research Group "Medical Signal Processing", Germany

ORGANIZING AND STEERING COMMITTEES

CONFERENCE CO-CHAIRS

Ana Fred, IST- Technical University of Lisbon, Portugal

Joaquim Filipe, INSTICC / Polytechnic Institute of Setúbal, Portugal

Hugo Gamboa, Telecommunications Institute / Portugal

Jorge Cardoso, University of Madeira - UMA / Madeira, Portugal

PROGRAM CO-CHAIRS

Luís Azevedo, IST- Technical University of Lisbon / Portugal

Ana Rita Londral, ANDITEC/ Portugal

LOCAL CHAIR

Paulo Sampaio, University of Madeira - UMA, Portugal

PROCEEDINGS PRODUCTION

Paulo Brito, INSTICC, Portugal

Marina Carvalho, INSTICC, Portugal

Helder Coelhas, INSTICC, Portugal

Vera Coelho, INSTICC, Portugal

Andreia Costa, INSTICC, Portugal

Bruno Encarnação, INSTICC, Portugal

Bárbara Lima, INSTICC, Portugal

Vitor Pedrosa, INSTICC, Portugal

Vera Rosário, INSTICC, Portugal

Mónica Saramago, INSTICC, Portugal

CD-ROM PRODUCTION

Paulo Brito, INSTICC, Portugal

GRAPHICS PRODUCTION AND WEB DESIGNER

Marina Carvalho, INSTICC, Portugal

SECRETARIAT AND WEBMASTER

Marina Carvalho, INSTICC, Portugal

PROGRAM COMMITTEE

Osman Abul, TOBB University of Economics and Technology, Turkey

Arnold Baca, University of Vienna, Austria

Iddo Bante, University of Twente, the Netherlands

Jyh-Horng Chen, National Taiwan University, Taiwan, Province of China

Kay Connelly, Indiana University, U.S.A.

Amar Das, Stanford University, U.S.A.

Adrie Dumay, TNO, the Netherlands

Eduardo B. Fernandez, Florida Atlantic University, U.S.A.

Alexandru Floares, Oncological Institute Cluj-Napoca, Romania

Jose Fonseca, Universidade Nova de Lisboa, Portugal

Toshio Fukuda, Nagoya University, Japan

M. Chris Gibbons, Johns Hopkins University, U.S.A.

David Greenhalgh, University of Strathclyde, U.K.

Jin-Kao Hao, University of Angers, France

Tin Ho, Bell Labs, Alcatel-Lucent, U.S.A.

John H. Holmes, University of Pennsylvania School of Medicine, U.S.A.

Chun-Hsi Huang, University of Connecticut, U.K.

Benjamin Jung, Empirica - Gesellschaft für Kommunikations - und Technologieforschung mbH, Germany

Stavros Karkanis, Technological Educational Institute of Lamia, Greece

Andreas Kerren, Växjö University, Sweden

Georgios Kontaxakis, Universidad Politécnica de Madrid, Spain

Nigel Lovell, University of New South Wales, Australia

Andrew Mason, University of Auckland, New Zealand

Boleslaw Mikolajczak, University of Massachusetts Dartmouth, U.S.A.

Ahmed Morsy, Cairo University, Egypt

Laura Roa, University of Seville, Spain

Jean Roberts, University of Central Lancashire, U.K.

Joel Rodrigues, Institue of Telecommunications / University of Beira Interior, Portugal

Marcos Rodrigues, Sheffield Hallam University, U.K.

George Sakellaropoulos, University of Patras, Greece

Meena Kishore Sakharkar, Nanyang Technologcail University, Singapore

Ovidio Salvetti, ISTI-CNR, Italy

Nickolas Sapidis, University of the Aegean, Greece

Sepe Sehati, Oxford Brookes University, U.K.

Boris Shishkov, IICREST / University of Twente, the Netherlands

Iryna Skrypnyk, University of Jyvaskyla, Finland

John Stankovic, University of Virginia, U.S.A.

Ron Summers, Loughborough University, U.K.

Adrian Tkacz, BioMedical-IT, Poland

Aristides Vagelatos, RACTI, Greece

Athanasios Vasilakos, University of Western Macedonia, Greece

Taieb Znati, University of Pittsburgh, U.S.A.

AUXILIARY REVIEWERS

Sara Colantonio, ISTI-CNR, Italy

Teduh Dirgahayu, University of Twente, the Netherlands

Ana Sofia Fernandes, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Portugal

Hailiang Mei, University of Twente, the Netherlands

Davide Moroni, ISTI-CNR, Italy

John Sarivougioukas, General Hospital of Athens "G. Gennimatas", Greece

SELECTED PAPERS BOOK

A number of selected papers presented at HEALTHINF 2008 will be published by Springer, in a book entitled Biomedical Engineering Systems and Technologies. This selection will be done by the conference Chair and program co-chairs, among the papers actually presented at the conference, based on a rigorous review by the BIOSTEC 2008 program committee members.

OFFICIAL CARRIER

FOREWORD

This volume contains the proceedings of the *First International Conference on Health Informatics* (HEALTHINF 2008), organized by the Institute for Systems and Technologies of Information Control and Communication (INSTICC) and the University of Madeira, technically co-sponsored by the IEEE Engineering in Medicine and Biology Society (EMB) and the Workflow Management Coalition (WfMC), in cooperation with AAAI.

The purpose of the *International Conference on Health Informatics* is to bring together researchers and practitioners interested in the application of information and communication technologies (ICT) to healthcare and medicine in general and to the specialized support to persons with special needs in particular.

Databases, networking, graphical interfaces, intelligent decision support systems and specialized programming languages are just a few of the technologies currently used in medical informatics. Mobility and ubiquity in healthcare systems, standardization of technologies and procedures, certification, privacy are some of the issues that medical informatics professionals and the ICT industry in general need to address in order to further promote ICT in healthcare. In the case of medical rehabilitation and assistive technology the use of ICT has had important results in the enhancement of the quality of life, contributing to a full integration of all citizens in the societies they are also part of. HEALTHINF is a forum for debating all these aspects. Furthermore, this conference is also a meeting place for those interested in understanding the human and social implications of technology, not only in healthcare systems but in other aspects of human-machine interaction such as accessibility issues.

HEALTHINF is one of three integrated conferences that are co-located and constitute the International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC). The other two component conferences are BIOSIGNALS (International Conference on Bio-inspired Systems and Signal Processing) and BIODEVICES (International Conference on Biomedical Electronics and Devices).

The joint conference, BIOSTEC, has received 494 paper submissions from more than 40 countries in all continents. 65 papers were published and presented as full papers, i.e. completed work (8 pages/30' oral presentation), 189 papers reflecting work-in-progress or position papers were accepted for short presentation, and another 86 contributions were accepted for poster presentation. These numbers, leading to a "full-paper" acceptance ratio below 14% and a total oral paper presentations acceptance ratio below 52%, show the intention of preserving a high quality forum for the next editions of this conference.

The conference included a panel and six invited talks delivered by internationally distinguished speakers, namely: Sergio Cerutti, Kevin Warwick, F. H. Lopes da Silva, Vipul Kashyap, David Hall and Albert Cook. Their participation has positively contributed to reinforce the overall quality of the Conference and to provide a deeper understanding of the field of Biomedical Engineering Systems and Technologies.

FOREWORD (CONT.)

The proceedings of the conference will be indexed by several major indices including DBLP, INSPEC and ISI-Proceedings and it will also be submitted for indexing to EI. A book with the revised versions of a short list of selected papers from the conference will be published by Springer-Verlag in the new CS book series: Communications in Computer and Information Science (CCIS). Additionally, a special issue of the IEEE Transactions on Biomedical Circuits and Systems will be edited based on the very best papers of the conference.

The program for this conference required the dedicated effort of many people. Firstly, we must thank the authors, whose research and development efforts are recorded here. Secondly, we thank the members of the program committee and the additional reviewers for their diligence and expert reviewing. Thirdly, we thank the keynote speakers for their invaluable contribution and for taking the time to synthesise and prepare their talks. Fourthly, we thank the program chairs, Luis Azevedo and Ana Rita Londral, whose collaboration was much appreciated. Finally, special thanks to all the members of the INSTICC team, especially Marina Carvalho at the conference secretariat, and the local organising committee from the University of Madeira, especially Jorge Cardoso and Paulo Sampaio, whose collaboration was fundamental for the success of this conference.

This year, the organization will distribute two paper awards at the conference closing session: the best paper award and the best student paper award. The decision was mainly based on the paper classifications provided by the Program Committee.

We wish you all an exciting conference and an unforgettable stay in the lovely island of Madeira. We hope to meet you again next year for the 2nd HEALTHINF, details of which are available at http://www.healthinf.org.

Joaquim Filipe

INSTICC/Polytechnic Institute of Setúbal

CONTENTS

INVITED SPEAKERS

KEYNOTE LECTURES

MULTIVARIATE, MULTIORGAN AND MULTISCALE INTEGRATION OF INFORMATION IN BIOMEDICAL SIGNAL PROCESSING

Sergio Cerutti
Department of Bioengineering, Polytechnic University, Milano, Italy
sergio.cerutti@polimi.it

Abstract: Biomedical signals carry important information about the behavior of the living systems under studying. A proper processing of these signals allows in many instances to obtain useful physiological and clinical information. Many advanced algorithms of signal and image processing have recently been introduced in such an advanced area of research and therefore important selective information is obtainable even in presence of strong sources of noise or low signal/noise ratio. Traditional stationary signal analysis together with innovative methods of investigation of dynamical properties of biological systems and signals in second-order or in higher-order approaches (i.e., in time-frequency, time-variant and time-scale analysis, as well as in non linear dynamics analysis) provide a wide variety of even complex processing tools for information enhancement procedures. Another important innovative aspect is also remarked: the integration between signal processing and modeling of the relevant biological systems is capable to directly attribute patho-physiological meaning to the parameters obtained from the processing and viceversa the modeling fitting could certainly be improved by taking into account the results from signal processing procedure. Such an integration process could comprehend parameters and observations detected at different scales, at different organs and with different modalities. This approach is reputed promising for obtaining an olistic view of the patient rather than an atomistic one which considers the whole as a simple sum of the single component parts.

BRIEF BIOGRAPHY

Sergio Cerutti is Professor in Biomedical Signal and Data Processing at the Department of Bioengineering of the Polytechnic University in Milano, Italy. In the period 2000-2006 he has been the Chairman of the same Department. His research interests are mainly in the following topics: biomedical signal processing (ECG, blood pressure signal and respiration, cardiovascular variability signals, EEG and evoked potentials), neurosciences and cardiovascular modelling. In his research activity he has put emphasis on the integration of information at different modalities, at different sources and at different scales in various physiological systems. Since 1983 he has taught a course at a graduate and a doc level on Biomedical Signal Processing and Modelling at Engineering Faculties (Milano and Roma) as well as at Specialisation Schools of Medical Faculties (Milano and Roma). He has been Elected Member of IEEE-EMBS AdCom (Region 8) in the period 1993-1996. He is actually Fellow Member of IEEE and of EAMBES and Associate Editor of IEEE Trans BME. He is a member of the Steering Committee of the IEEE-EMBS Summer School on Biomedical Signal Processing: he was the local organiser of four Summer Schools held in Siena. He has been Visiting Professor at Harvard-MIT Division Health Science and Technology, Boston, USA for an overall period of 1 year. He is the Author of more than 400 international scientific contributions (more than 180 on indexed scientific journals).

1 INTRODUCTION

Biomedical signals and imaging carry important information about the behavior of the living systems under studying. A proper processing of these signals and images allow in many instances to obtain useful physiological and clinical information. Actually, many advanced algorithms of digital signal and image processing are at disposal and therefore

important selective information is now obtainable even in presence of strong sources of noise or low signal/noise ratio. In most of the cases it is not sure whether such sources might derive even by complex and unknown interactions with other biological systems whose implications could be important from the physiological or clinical standpoints. Traditional stationary signal analysis together with innovative methods of investigation of dynamical properties of biological systems and signals in second-order or in higher-order approaches (i.e., in time-frequency, time-variant and time-scale analysis, as well as in non linear dynamics analysis) provide a wide variety of even complex processing tools for information enhancement procedures in the challenging studying of a better explanation of many physiological and clinical phenomena.

2 INTEGRATION BETWEEN SIGNAL PROCESSING AND PHYSIOLOGICAL MODELING

Another important innovative aspect to improve the information content from biomedical data is constituted by the integration between signal processing and modeling of the relevant biological systems, thus directly attributing patho-physiological meaning to the model parameters obtained from the processing; and, viceversa, the modeling fitting could certainly be improved by taking into account the results from signal/image processing procedures.

3 MONOVARIATE AND MULTIVARIATE SIGNAL PROCESSING

Other kinds of integration may be fulfilled, taking into account more signals from the same system in a multivariate way (i.e. from a single-lead vs multichannel EEG or ECG analysis) and combining also the action of different systems such as autonomic nervous system, cardiovascular and respiratory systems, etc. Sleep is a formidable example of multiorgan involvement in both physiological (sleep staging and correlation with cardiorespiratory system) and pathological conditions (sleep apnea, sleep deprivation, restless leg syndrome and so on).

4 MULTISCALE APPROACH

Further, modern rehabilition techniques (motor and /or cognitive) make use actually of objective indices obtained from the patient's biosignals and images to better "personalize" rehabilitation protocols (from EEG, EP's, ERP's, MRI, fMRI, NIRS, etc). In neurosciences such an integration process could comprehend parameters and observations detected also at different scales, from genome and proteome up to the single organ and to the entire body compartment. Examples will be described where an animal model (murine model) is developed by alterating a gene putative to a determined pathology (i.e.epilepsy) and changes in EEG signals are studied (spike/wave occurrences and modifications in signal power bands). In clinical applications, it is worth mentioning the important data fusion which could be fulfilled by the integration of simultaneous EEG recordings and fMRI in some epileptic patients during inter-critical or critical events.

Finally, another important integration can be obtained along different observation scales. Traditionally, biological signal analysis is carried out at the level of organ or system to be investigated (i.e., ECG or EEG signal, arterial blood pressure, respiration and so on). It is very clear the advantage of correlating this information with that one obtained about the same system, but at different scale level, i.e. at cellular level or even at subcellular level (for example, analyzing possible genetic correlates or typical patterns of proteins or even DNA/RNA sequences). Biomedical engineering as a dedicated discipline may strongly contribute to this multiscale information processing

Along this approach line, even the long-QT syndrome, can be efficiently studied at different scale level: a mutation in a portion of gene SCN5A which presents a phenotype compatible to long-QT3 type, is known to produce an altered function of Na+ channels. Through a proper model which describes the functioning of ventricular cells is possible to evidence that this alteration may induce a prolongation of QT duration, as detected on ECG tracing. This event is further correlated with an increased risk of ventricular tachyarrhythmias. Hence, the path is completed: from the genetic expression up to the disease manifestation (Clancy and Rudy, 1999), (Priori ey al., 2003). Many different signal processing and modeling are involved in this paradigmatic example: an integration along the various scales of observation may undoubtedly contribute to a better

understanding of the complex pathophysiological correlates.

A great effort is on course nowadays for creating very large databases and networking of models and technologies for integrating such information (Physiome project (Hunter et al., 2002), (Rudy, 2000) to be connected with Genome and Proteome projects and Virtual Physiological Human project – VPH – which is inserted into the activities of the 7th Framework Programme of EU).

Other examples are constituted by the studying of the profile of expressed proteins in 2D-gel supports, or after mass-spectrometry analysis, relative to a variety of pathologies (i.e. epilepsy, peripheral neuropathies or Amyotrophic Lateral Sclerosis (ALS), or in oncological studies) thus singling out the set of proteins which present a correlate with the pathology in respect to the control group.

This overall approach is reputed promising for obtaining an olistic view of the patient rather than an atomistic one which considers the whole as a simple sum of the single component parts.

REFERENCES

G. Baselli, S. Cerutti, S. Civardi, A. Malliani and M. Pagani, Cardiovascular variability signals: towards the identification of a closed-loop model of the neural control mechanisms, *IEEE Trans Biomed Eng*., vol.35(12):1033-46(1988).

E. N. Bruce, *Biomedical Signal Processing and Signal Modelling*, J. Wiley (2001).

C. E. Clancy and Y. Rudy, Linking a genetic defect to its cellular phenotype in a cardiac arrhythmia, *Nature*, vol. 400(6744):566-569 (1999).

A. Cohen, *Biomedical Signal Analysis*, Vol. I, II, CRC Press (1986).

P. Hunter, P. Robbins and D. Noble, The IUPS human physiome project, *Pflugers Arch-Eur. J. Physiol*., vol. 445:1-9 (2002).

R. I. Kitney, A nonlinear model for studing oscillations in the blood pressure control system*, J. Biomed. Eng*., vol I,n. 2: 88-89 (1979).

H. P. Koepchen, History of studies and concepts of blood pressure wave, in: *Mechanisms of blood pressure waves (*Miyakawa K., Koepchen, H.P., Polosa C. eds): 3-23, Springer-Verlag (1984).

M. C. Mackey and J.C. Milton, Dynamical diseases, *Ann N Y Acad.Sc.,* vol.504:16-32 (1987).

C.-K. Peng, S. Havlin, J.M. Hausdorff, J.E. Mietus, H.E. Stanley and A.L. Goldberger, Fractal mechanisms and heart rate dynamics long-range correlation and their breakdown with diseases, *Journal of Electrocardiology*, vol.28: 59-65 (1995).

S. G. Priori, P.J. Schwartz, C. Napolitano, R. Bloise, E. Ronchetti, M. Grillo, A. Vicentini, C. Spazzolini, J. Nastoli, G. Bottelli, R. Folli and D. Cappelletti, Risk stratification in the long-QT sindrome, *N Engl J Med*., May 8, vol.348(19):1866-74 (2003).

Y. Rudy, From Genome to Physiome: integrative models of cardiac excitation, *Ann. Biom. Engineer*, vol 28: 945-950 (2000).

Cerutti S, Esposti F, Ferrario M, Sassi R, Signorini MG, Long-term invariant parameters obtained from 24-h Holter recordings: a comparison between different analysis techniques, Chaos, vol.17(1):015108, (2007).

Foffani G, Ardolino G, Rampini P, Tamma F, Caputo E, Egidi M, Cerutti S, Barbieri S, Priori A., Physiological recordings from electrodes implanted in the basal ganglia for deep brain stimulation in Parkinson's disease: the relevance of fast subthalamic rhythms, Acta Neurochir Suppl, vol. 93:97-9, (2005).

OUTTHINKING AND ENHANCING BIOLOGICAL BRAINS

Kevin Warwick
University of Reading, UK

Keywords: Brain-Computer Interface, Biological systems, Implant technology, Feedback control.

Abstract: In this paper an attempt has been made to take a look at how the use of implant and electrode technology can now be employed to create biological brains for robots, to enable human enhancement and to diminish the effects of certain neural illnesses. In all cases the end result is to increase the range of abilities of the recipients. An indication is given of a number of areas in which such technology has already had a profound effect, a key element being the need for a clear interface linking the human brain directly with a computer. An overview of some of the latest developments in the field of Brain to Computer Interfacing is also given in order to assess advantages and disadvantages. The emphasis is clearly placed on practical studies that have been and are being undertaken and reported on, as opposed to those speculated, simulated or proposed as future projects. Related areas are discussed briefly only in the context of their contribution to the studies being undertaken. The area of focus is notably the use of invasive implant technology, where a connection is made directly with the cerebral cortex and/or nervous system.

Tests and experimentation which do not involve human subjects are invariably carried out *a priori* to indicate the eventual possibilities before human subjects are themselves involved. Some of the more pertinent animal studies from this area are discussed including our own involving neural growth. The paper goes on to describe human experimentation, in which neural implants have linked the human nervous system bi-directionally with technology and the internet. A view is taken as to the prospects for the future for this implantable computing in terms of both therapy and enhancement.

BRIEF BIOGRAPHY

Kevin Warwick is Professor of Cybernetics at the University of Reading, England, where he carries out research in artificial intelligence, control, robotics and cyborgs. He is also Director of the University KTP Centre, which links the University with Small to Medium Enterprises and raises £2.5 million each year in research income. As well as publishing 500 research papers, Kevin is perhaps best known for his experiments into implant technology. He has been awarded higher doctorates (DScs) both by Imperial College and the Czech Academy of Sciences, Prague. He was presented with The Future of Health Technology Award in MIT, was made an Honorary Member of the Academy of Sciences, St. Petersburg and in 2004 received The IEE Achievement Medal.

1 INTRODUCTION

Research is being carried out in which biological signals of some form are measured, are acted upon by some appropriate signal processing technique and are then employed either to control a device or as an input to some feedback mechanism (Penny et al., 2000), (Roitberg, 2005). In many cases neural signals are employed, for example Electroencephalogram (EEG) signals can be measured externally to the body, using externally adhered electrodes on the scalp (Wolpaw et al., 1990) and can then employed as a control input. Most likely this is because the procedure is relatively simple from a research point of view and is not particularly taxing on the researchers involved. However, reliable interpretation of EEG data is extremely complex – partly due to both the compound nature of the multi-neuronal signals being measured and the difficulties in recording such highly attenuated

In the last few years interest has also grown in the use of real-time functional Magnetic Resonance

Imaging (fMRI) for applications such as computer cursor control. This typically involves an individual activating their brain in different areas by reproducible thoughts (Warwick, 2007) or by recreating events (Pan et al., 2007). Alternatively fMRI and EEG technologies can be combined so that individuals can learn how to regulate Slow Cortical Potentials (SCPs) in order to activate external devices (Hinterberger et al., 2005). Once again the technology is external to the body. It is though relatively expensive and cumbersome.

It is worth noting that external monitoring of neural signals, by means of either EEG analysis or indeed fMRI, leaves much to be desired. Almost surely the measuring technique considerably restricts the user's mobility and, as is especially the case with fMRI, the situation far from presents a natural or comfortable setting. Such systems also tend to be relatively slow, partly because of the nature of recordings via the indirect connection, but also because it takes time for the individual themselves to actually initiate changes in the signal. As a result of this, distractions, both conscious and sub-conscious, can result in false indicators thus preventing the use of such techniques for safety critical, highly dynamic and, to be honest, most realistic practical applications. Despite this, the method can enable some individuals who otherwise have extremely limited communication abilities to operate some local technology in their environment, and, in any case, it can serve as a test bed for a more direct and useful connection.

The definition of what constitutes a Brain-Computer Interface (BCI) is extremely broad. A standard keyboard could be so regarded. It is clear however that various wearable computer techniques and virtual reality systems, e.g. glasses containing a miniature computer screen for a remote visual experience (Mann, 1997), are felt by some researchers to fit this category. Although it is acknowledged that certain body conditions, such as stress or alertness, can be monitored in this way, the focus of this paper is on bidirectional BCIs and is more concerned with a direct connection between a biological brain and technology, and ultimately a human and technology.

2 *IN VIVO* STUDIES

Non-human animal studies can be considered to be a pointer for what is potentially achievable with humans in the future. As an example, in one particular animal study the extracted brain of a lamprey, retained in a solution, was used to control the movement of a small wheeled robot to which it was attached (Reger et al., 2000). The lamprey innately exhibits a response to light reflections on the surface of water by trying to align its body with respect to the light source. When connected into the robot body, this response was utilised by surrounding the robot with a ring of lights. As different lights were switched on and off, so the robot moved around its corral, trying to position itself appropriately.

Meanwhile in studies involving rats, a group of rats were taught to pull a lever in order to receive a suitable reward. Electrodes were then chronically implanted into the rats' brains such that the reward was proffered when each rat thought (one supposes) about pulling the lever, but before any actual physical movement occurred. Over a period of days, four of the six rats involved in the experiment learned that they did not in fact need to initiate any action in order to obtain a reward; merely thinking about it was sufficient (Chapin, 2004).

In another series of experiments, implants consisting of microelectrode arrays have been positioned into the frontal and parietal lobes of the brains of two female rhesus macaque monkeys. Each monkey learned firstly how to control a remote robot arm through arm movements coupled with visual feedback, and it is reported that ultimately one of the monkeys was able to control the arm using only brain derived neural signals with no associated physical movement. Notably, control signals for the reaching and grasping movements of the robotic arm were derived from the same set of implanted electrodes (Carmena et al., 2003), (Nicolelis et al., 2000).

Such promising results from animal studies have given the drive towards human applications a new impetus.

3 ROBOT WITH A BIOLOGICAL BRAIN

Human concepts of a robot may involve a little wheeled device, perhaps a metallic head that looks roughly human-like or possibly a biped walking robot. Whatever the physical appearance our idea tends to be that the robot might be operated remotely by a human, or is being controlled by a simple programme, or even may be able to learn with a

microprocessor/computer as its brain. We regard a robot as a machine.
In a present project neurons are being cultured in a laboratory in Reading University to grow on and interact with a flat multi-electrode array. The neural culture, a biological brain, can be electronically stimulated via the electrodes and its trained response can be witnessed.

The project now involves networking the biological brain to be part of a robot device. In the first instance this will be a small wheeled robot. The input (sensory) signals in this case will be only the signals obtained from the wheeled robot's ultrasonic sensors. The output from the biological brain will be used to drive the robot around. The goal of the project initially will be to train the brain to drive the robot forwards without bumping into any object. Secondly, a separate biological brain will be grown to be the thinking process within a robot head (called Morgui) which houses 5 separate sensory inputs.

What this means is that the brain of these robots will shortly be a biological brain, not a computer. All the brain will know is what it perceives from the robot body and all it will do will be to drive the robot body around or control the robot head respectively. The biological brain will, to all intents and purposes, be the brain of the robot. It will have no life, no existence outside its robotic embodiment.

Clearly this research alters our concept of what a robot is, particularly in terms of ethical and responsibility issues. If a role of animal research is to open up possibilities for future human trials, then in this case the research could well be opening a window on the ultimate possibility of human neurons being employed in a robot body. All the 'human' brain would know would be its life as a robot.

4 HUMAN APPLICATION

At the present time the general class of Brain-Computer Interfaces (BCIs) for humans, of one form or another, have been specifically developed for a range of applications including military weapon and drive systems, personnel monitoring and for games consoles. However, by far the largest driving force for BCI research to date has been the requirement for new therapeutic devices such as neural prostheses.

The most ubiquitous sensory neural prosthesis in humans is by far the cochlea implant (Fin and LoPresti, 2003). Here the destruction of inner ear hair cells and the related degeneration of auditory nerve fibres results in sensorineural hearing loss. As such, the prosthesis is designed to elicit patterns of neural activity via an array of electrodes implanted into the patient's cochlea, the result being to mimic the workings of a normal ear over a range of frequencies. It is claimed that some current devices restore up to approximately 80% of normal hearing, although for most recipients it is sufficient that they can communicate to a respectable degree without the need for any form of lip reading. The typically modest success of cochlea implantation is related to the ratio of stimulation channels to active sensor channels in a fully functioning ear. Recent devices consist of up to 32 channels, whilst the human ear utilises upwards of 30,000 fibres on the auditory nerve. There are now reportedly well over 10,000 of these prostheses in regular operation.

Studies investigating the integration of technology with the human central nervous system have varied from merely diagnostic to the amelioration of symptoms (Warwick and Gasson, 2004). In the last few years some of the most widely reported research involving human subjects is that based on the development of an artificial retina (Rizzo, 2001). Here, small electrode arrays have been successfully implanted into a functioning optic nerve. With direct stimulation of the nerve it has been possible for the otherwise blind recipient to perceive simple shapes and letters. The difficulties with restoring sight are though several orders of magnitude greater than those of the cochlea implant simply because the retina contains millions of photodetectors that need to be artificially replicated. An alternative is to bypass the optic nerve altogether and use cortical surface or intracortical stimulation to generate phosphenes (Dobelle, 2000).

Most invasive BCIs monitor multi-neuronal intracortical action potentials, requiring an interface which includes sufficient processing in order to relate recorded neural signals with movement intent. Problems incurred are the need to position electrodes as close as possible to the source of signals, the need for long term reliability and stability of interface in both a mechanical and a chemical sense, and adaptivity in signal processing to deal with technological and neuronal time dependence. However, in recent years a number of different collective assemblies of microelectrodes have been successfully employed both for recording and stimulating neural activity. Although themselves of small scale, nevertheless high density

connectors/transmitters are required to shift the signals to/from significant signal processing and conditioning devices and also for onward/receptive signal transmission.

Some research has focussed on patients who have suffered a stroke resulting in paralysis. The most relevant to this paper is the use of a '3rd generation' brain implant which enables a physically incapable brainstem stroke victim to control the movement of a cursor on a computer screen (Kennedy, 2000), (Kennedy, 2004). Functional Magnetic Resonance Imaging (fMRI) of the subject's brain was initially carried out to localise where activity was most pronounced whilst the subject was thinking about various movements. A hollow glass electrode cone containing two gold wires and a neurotrophic compound (giving it the title 'Neurotrophic Electrode') was then implanted into the motor cortex, in the area of maximum activity. The neurotrophic compound encouraged nerve tissue to grow into the glass cone such that when the patient thought about moving his hand, the subsequent activity was detected by the electrode, then amplified and transmitted by a radio link to a computer where the signals were translated into control signals to bring about movement of the cursor. With two electrodes in place, the subject successfully learnt to move the cursor around by thinking about different movements. Eventually the patient reached a level of control where no abstraction was needed – to move the cursor he simply thought about moving the cursor. Notably, during the period that the implant was in place, no rejection of the implant was observed; indeed the neurons growing into the electrode allowed for stable long-term recordings.

Electronic neural stimulation has proved to be extremely successful in other areas, including applications such as the treatment of Parkinson's disease symptoms. With Parkinson's Disease diminished levels of the neurotransmitter dopamine cause over-activation in the ventral posterior nucleus and the subthalamic nucleus, resulting in slowness, stiffness, gait difficulties and hand tremors. By implanting electrodes into the subthalamic nucleus to provide a constant stimulation pulse, the over activity can be inhibited allowing the patient, to all external intents and purposes, to function normally (Pinter et al., 1999).

5 BRAIN WITHIN A BRAIN

Ongoing research, funded by the UK Medical Research Council, is investigating how the onset of tremors can be accurately predicted such that merely a stimulation current burst is required rather than a constant pulsing (Gasson et al., 2005: pp.16/1-16/4). This has implications for battery inter-recharge periods as well as limiting the extent of in-body intrusive signalling. The deep brain stimulator can be used to collect local field potential (LFP) signals generated by the neurons around the deep brain electrodes (Gasson et al., 2005: pp.16/1-16/4). Determining the onset of events can be investigated by using fourier transforms to transfer the time based signal to a frequency based spectrogram to determine the change in frequency at the critical time period. However, in addition to that, the frequency changes in the period of time immediately prior to the tremor occurrence can give important information.

Fig.1 shows the results of an initial attempt to train an artificial neural network to indicate not only that a Parkinsonian tremor is present but also that one is very likely to occur in the near future. The aim of this research is that, once a reliable predictor has been obtained, the stimulating pulsing will only be enacted when a tremor is predicted, in order to stop the actual physical tremor occurring before it even starts in the first place.

The bottom trace in Fig.1 shows emg (muscular) signals, measured externally, associated with movement due to the tremors. It can be seen that the tremors in this incident actually start at around the 45 to 50 second point. The trace just above this indicates the corresponding electrical data measured as deep brain Local Field Potentials in the Sub-Thalamic Nucleus of the patient involved. It can be witnessed how, in this case, the electrical data takes on a different form (in terms of variance at least) at around the 45 to 50 second point. The four top plots meanwhile indicate the outputs from 4 differently structured artificial neural networks, based on multi-layer perceptrons with different numbers of neurons in the hidden (middle) layer.

It can be seen how, for each network, the output of the network goes high (logic 1) at the 45 to 50 second point, to indicate the presence of a Parkinsonian tremor. This is all well and good, what is important however is that the output of the networks also briefly goes high around the 30

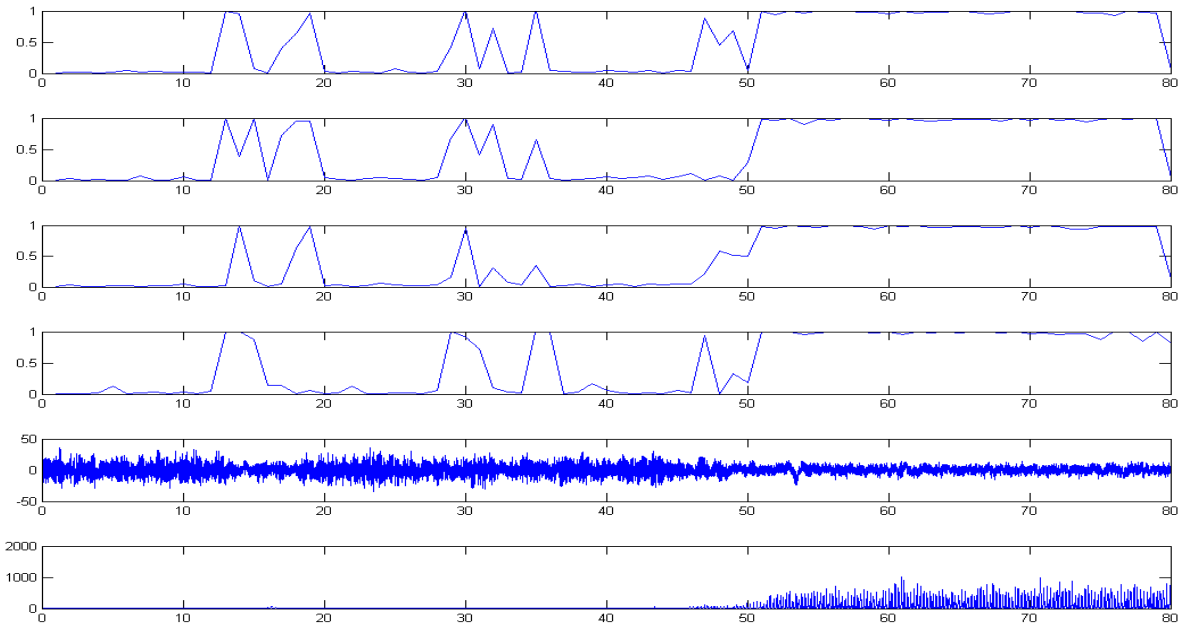


Figure 1: Time plot of the onset of a Parkisonian tremor incident with corresponding artificial neural network indicators.

second point and this can be seen as an indication of the fact that a tremor will shortly occur. Ongoing research is involved with selection of the type and number of inputs to the network, presently these being based on the energy spectrum in different frequency ranges. The networks are also being tested on considerable amounts of resting data, that is long periods of brain activity where no tremors at all actually occur in patients. Clearly the aim is that a network will not give false predictions of tremors.

In fact false positive predictions are not so much of a critical problem. The end result with a false positive is that a stimulation may occur when it is not strictly necessary. In any event no actual tremor would occur, which is indeed a good outcome, however unnecessary energy would have been used – in fact if numerous false predictions occurred the intelligent stimulator would tend toward the present 'blind' stimulator. Effectively the occasional false positive prediction is perhaps not a problem, unless it became a regular occurrence. The good news is that results show that the network can be readily tuned to avoid false positives anyway.

6 GENERAL IMPLANT STUDIES

Some of the most impressive human research to date has been carried out using the microelectrode array, shown in Figure 2. The individual electrodes are only 1.5mm long and taper to a tip diameter of less than 90 microns. Although a number of trials not using humans as a test subject have occurred (Branner and Normann, 2000), human tests are at present limited to two studies. In the second of these the array has been employed in a recording only role (Donoghue et al., 2002), (Donoghue et al., 2004), (Friehs et al., 2004), most notably recently as part of the 'Braingate' system. Essentially activity from a few neurons monitored by the array electrodes is decoded into a signal to direct cursor movement. This has enabled an individual to position a cursor on a computer screen, using neural signals for control combined with visual feedback. The first use of the microelectrode array (Figure 2) will be discussed in the following section as this has considerably broader implications which extend the capabilities of the human recipient.

A key selection point at the present time are what type of implant to employ, as several different possibilities exist, ranging from single electrode devices to multielectrode needles which contain electrode points at different depths to multielectrode arrays which either contain a number of electrodes which penetrate to the same depth (as in Figure 2) or are positioned in a banked/sloped arrangement. A further key area of consideration is the exact positioning of a BCI. In particular certain areas of the brain are, apparently, only really useful for monitoring purposes whilst others are more useful for stimulation.

Actually deriving a reliable command signal from a collection of captured neural signals is not necessarily a simple task, partly due to the complexity of signals recorded and partly due to time constraints in dealing with the data. In some cases however it can be relatively easy to look for and obtain a system response to certain anticipated neural signals – especially when an individual has trained extensively with the system. In fact neural signal shape, magnitude and waveform with respect to time are considerably different to the other signals that it is possible to measure in this situation.

If a greater understanding is required of neural signals recorded, before significant progress can be made, then this will almost surely present a major problem. This is especially true if a number of simultaneous channels are being employed, each requiring a rate of digitization of (most likely) greater than 20KHz in the presence of unwanted noise. For real time use this data will also need to be processed within a few milliseconds (100 milliseconds at most). Further, although many studies have looked into the extraction of command signals (indicating intent) from measured values, it is clear that the range of neural activity is considerable. Even in the motor area not only are motor signals present but so too are sensory, cognitive, perceptual along with other signals, the exact purpose of which is not clear – merely classifying them as noise is not really sufficient and indeed can be problematic when they are repeated and apparently linked in some way to activity.

It is worth stressing here that the human brain and spinal cord are linking structures, the functioning of which can be changed through electronic stimulation such as that provided via an electrode arrangement. This type of technology therefore offers a variety of therapeutic possibilities. In particular the use of implanted systems when applied to spinal cord injured patients, in whom nerve function is disordered, was described in (Warwick, 2004) as having the following potential benefits (among others):

1. Re-education of the brain and spinal cord through repeated stimulation patterns
2. Prevention of spinal deformity
3. Treatment of intractable neurogenic and other pain
4. Assisting bladder emptying
5. Improving bowel function
6. Treatment of spasticity
7. Improvement of respiratory function – assisting coughing and breathing
8. Reduction of cardiovascular maleffects
9. Prevention of pressure sores – possibly providing sensory feedback from denervated areas
10. Improvement and restoration of sexual function
11. Improved mobility
12. Improved capability in daily living, especially through improved hand, upper limb and truncal control

Sensate prosthetics is another growing application area of neural interface technology, whereby a measure of sensation is restored using signals from small tactile transducers distributed within an artificial limb (Fin and LoPresti, 2003). The transducer output can be employed to stimulate the sensory axons remaining in the residual limb which are naturally associated with a sensation. This more closely replicates stimuli in the original sensory modality, rather than forming a type of feedback using neural pathways not normally associated with the information being fed back. As a result it is supposed that the user can employ lower level reflexes that exist within the central nervous system, making control of the prosthesis more subconscious.

One final noteworthy therapeutic procedure is Functional Electrical Stimulation (FES), although it is debatable if it can be truly referred to as a BCI, however it aims to bring about muscular excitation, thereby enabling the controlled movement of limbs. FES has been shown to be successful for artificial hand grasping and release and for standing and walking in quadriplegic and paraplegic individuals as well as restoring some basic body functions such as bladder and bowel control (Grill and Kirsch, 2000). It must be noted though that controlling and coordinating concerted muscle movements for complex and generic tasks such as picking up an arbitrary object is proving to be a difficult, if not insurmountable, challenge.

In the cases described in which human subjects are involved, the aim on each occasion is to either restore functions since the individual has a physical problem of some kind or it is to give a new ability to an individual who has very limited motor abilities. In this latter case whilst the procedure can be regarded as having a therapeutic purpose, it is quite possible to provide an individual with an ability that they have in fact never experienced before. On the one hand it may be that whilst the individual in question has never previously experienced such an ability, some or most other humans have – in this

case it could be considered that the therapy is bringing the individual more in line with the "norm" of human abilities.

It is though also potentially possible to give extra capabilities to a human, to enable them to achieve a broader range of skills – to go beyond the "norm". Apart from the, potentially insurmountable, problem of universally deciding on what constitutes the "norm", extending the concept of therapy to include endowing an individual with abilities that allow them to do things that a perfectly able human cannot do raises enormous ethical issues. Indeed it could be considered that a cochlea implant with a wider frequency response range does just that for an individual or rather an individual who can control the curser on a computer screen directly from neural signals falls into this category. But the possibilities of enhancement are enormous. In the next section we consider how far things could be taken, by referring to relevant experimental results.

7 HUMAN ENHANCEMENT

The interface through which a user interacts with technology provides a distinct layer of separation between what the user wants the machine to do, and what it actually does. This separation imposes a considerable cognitive load upon the user that is directly proportional to the level of difficulty experienced. The main issue it appears is interfacing the human motor and sensory channels with the technology. One solution is to avoid this sensorimotor bottleneck altogether by interfacing directly with the human nervous system. It is certainly worthwhile considering what may potentially be gained from such an invasive undertaking.

Advantages of machine intelligence are for example rapid and highly accurate mathematical abilities in terms of 'number crunching', a high speed, almost infinite, internet knowledge base, and accurate long term memory. Additionally, it is widely acknowledged that humans have only five senses that we know of, whereas machines offer a view of the world which includes infra-red, ultraviolet and ultrasonic. Humans are also limited in that they can only visualise and understand the world around them in terms of a limited dimensional perception, whereas computers are quite capable of dealing with hundreds of dimensions. Also, the human means of communication, essentially transferring an electro-chemical signal from one brain to another via an intermediate, often mechanical medium, is extremely poor, particularly in terms of speed, power and precision. It is clear that connecting a human brain, by means of an implant, with a computer network could in the long term open up the distinct advantages of machine intelligence, communication and sensing abilities to the implanted individual.

As a step towards this more broader concept of human-machine symbiosis, in the first study of its kind, the microelectrode array (as shown in Figure 2) was implanted into the median nerve fibres of a healthy human individual (myself) in order to test *bidirectional* functionality in a series of experiments. A stimulation current direct onto the nervous system allowed information to be sent to the user, while control signals were decoded from neural activity in the region of the electrodes (Gasson et al., 2005:pp 365-375), (Warwick et al., 2003).

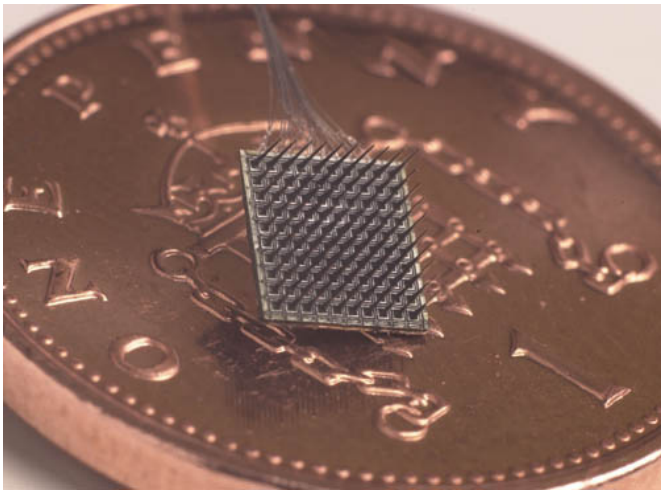


Figure 2: A 100 electrode, 4X4mm Microelectrode Array, shown on a UK 1 pence piece for scale.

In this way a number of experimental trials were successfully concluded (Warwick et al., 2004), (Warwick et al., 2005): In particular:

1. Extra sensory (ultrasonic) input was successfully implemented and made use of.
2. Extended control of a robotic hand across the internet was achieved, with feedback from the robotic fingertips being sent back as neural stimulation to give a sense of force being applied to an object (this was achieved between New York (USA) and Reading(UK))
3. A primitive form of telegraphic communication directly between the nervous systems of two humans was performed.
4. A wheelchair was successfully driven around by means of neural signals.
5. The colour of jewellery was changed as a result of neural signals – as indeed was the behaviour of a collection of small robots.

In each of the above cases it could be regarded that the trial proved useful for purely therapeutic reasons, e.g. the ultrasonic sense could be useful for an individual who is blind or the telegraphic communication could be very useful for those with certain forms of Motor Neurone Disease. However each trial can also be seen as a potential form of augmentation or enhancement for an individual. The question then arises as to how far should things be taken? Clearly enhancement by means of BCIs opens up all sorts of new technological and intellectual opportunities, however it also throws up a raft of different ethical considerations that need to be addressed directly.

8 ON STIMULATION

After extensive experimentation it was found that injecting currents below 80µA onto the median nerve fibers had little perceivable effect. Between 80µA and 100µA all the functional electrodes were able to produce a recognizable stimulation, with an applied voltage of 40 to 50 volts, dependant on the series electrode impedance. Increasing the current above 100µA had no apparent additional effect; the stimulation switching mechanisms in the median nerve fascicle exhibited a non-linear thresholding characteristic.

During this experimental phase, it was pseudo randomly decided whether a stimulation pulse was applied or not. The volunteer (myself), wearing a blindfold, was unaware of whether a pulse had been applied or not, other than by means of its effect in terms of neural stimulation. The user's accuracy in distinguishing between an actual pulse and no pulse at a range of amplitudes is shown in Figure 3.

In all subsequent successful trials, the current was applied as a bi-phasic signal with pulse duration of 200 µsec and an inter-phase delay of 100 µsec. A typical stimulation waveform of constant current being applied to one of the MEA's implanted electrodes is shown in Fig 4.

It was, in this way, possible to create alternative sensations via this new input route to the nervous system. Of the 5 enhancement features mentioned in the previous section, this one will be described, as an example, in further detail. Background information on the other enhancements can be found in a number of references, e.g. (Gasson et al., 2005:pp 365-375), (Warwick et al., 2003), (Warwick et al., 2004), (Warwick and Gasson, 2004).

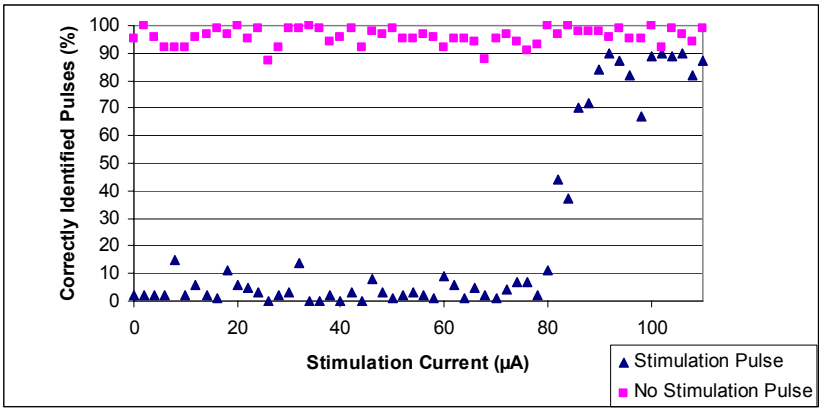


Figure 3: Effect of stimulation amplitude on the number of correctly identified pulses and absence of pulses (over 100 trials).

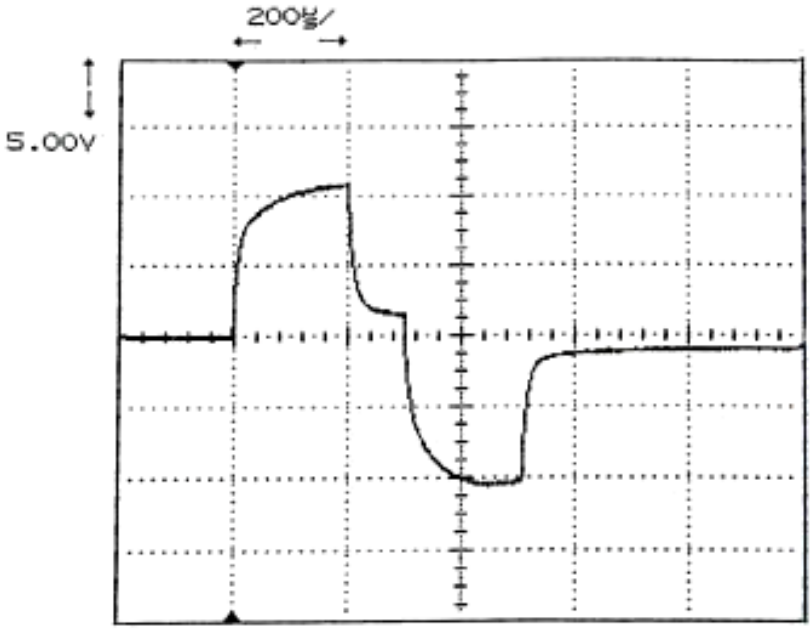


Figure 4: Voltage profile during one bi-phasic stimulation pulse cycle with a constant current of 80µA.

It must be reported that it took 6 weeks for my brain to repetitively recognize the stimulating signals accurately. This time period can be due to a number of contributing factors:

(a) The team had to learn which signals (what amplitude, frequency etc.) would be best in order to bring about a recognizable stimulation.

(b) The recipient's brain had to learn to recognize the new signals it was receiving.

(c) The bond between the recipient's nervous system and the implant was physically changing (becoming stronger).

9 EXTRA SENSORY EXPERIMENT

An experiment was set up to determine if the human brain is able to understand and successfully operate with sensory information to which it had not previously been exposed. Whilst it is quite possible to feed in such sensory information via a normal human sensory route, e.g. electromagnetic radar or infra-red signals are converted to visual, what we were interested in was feeding such signals directly onto the human nervous system, thereby bi-passing the normal human sensory input.

Ultrasonic sensors were fitted to the rim of a baseball cap (see Figure 5) and the output from these sensors, in the form of a proportional count, was employed to bring about a direct stimulation of the nervous system. Hence when no objects were in the vicinity of the sensors, no stimulation occurred, and as an object moved close by so the rate of stimulation pulses being applied increased in a linear fashion up to a pre-selected maximum rate. No increase in stimulation occurred when an object moved closer than 10cm to the sensors.

The ultrasonic sensors were open type piezoelectric ceramic transducers with conical metal resonators and operated at 40 KHz. These were used in a pair, one for transmit and one for receive, to give maximum sensitivity for small and distant objects. The most useful range for the experimentation was found to be 2 – 3m, this being also dependent on the size of object. A simple microcontroller was programmed to perform the echo ranging on the pair of transducers, and provide the range to the first detectable object only. This was translated into a stimulation pulse train, which operated on a single pin of the electrode array. Pins on the array had been tested for their suitability for stimulation by the earlier experimentation in which the recipient identified the presence or absence of stimulation pulse trains at various amplitudes and repetition frequencies.

Figure 5: Experimentation and testing of the ultrasonic baseball cap.

It was found that very little learning was required for the new ultrasonic sense to be used effectively and successfully – merely a matter of 5/6 minutes. This said it must be remembered that it had already taken several weeks for the recipient's brain to successfully, accurately recognize the current signals being injected.

As a result, in a witnessed experiment, the recipient, whilst wearing a blindfold, was able to move around successfully within a cluttered laboratory environment, albeit at a slower than normal walking pace. The sensory input was "felt" as a new form of sensory input (not as touch or movement) in the sense that the brain made a direct link between the signals being witnessed and the fact that these corresponded in a linear fashion to a nearby object.

10 CONCLUSIONS

External input-output interfaces with human and animal brains have been studied for many years. These are sometimes referred to as Brain-Computer Interfaces (BCIs) even though the interface may be external to the (human) body and its sensorimotor mechanism. In this paper an attempt has been made to put such systems in perspective. Emphasis has been placed on such interfaces that can be obtained by means of implanted devices through invasive surgery and actual direct neural connections. In particular a number of trials in this area have clearly shown the possibilities of monitoring, stimulating and enhancing brain functioning.

Although there is no distinct dividing line it is quite possible as far as humans are concerned to investigate BCIs in terms of those employed for direct therapeutic means and those which can have an enhanced role to play. It is clear that the interaction of electronic signals with the human brain can cause the brain to operate in a distinctly different manner. Such is the situation with the stimulator implants that are successfully used to counteract, purely electronically, the tremor effects associated with Parkinson's disease. Such technology can though potentially be employed to modify the normal functioning of the human brain and nervous system in a number of different ways.

The same stimulator, with slightly different positioning, has been shown to elicit feelings of sadness or happiness in the recipient. Given the nature of the intelligent stimulator described here it would appear to be possible to monitor, in real time, a human brain with a computer brain, and for the computer brain to predict when the human is going to feel sad – quite some time before they actually feel sad. In theory a signal could then be injected at

that time to make them feel happy, or at least to stop them actually ever feeling sad in the first place. Maybe this could be regarded as an electronic anti-depressant. There are of course questions about recreational use here – but this would need a deep brain implant which might well prove to be rather too onerous for most people.

Perhaps understandably, invasive BCIs are presently far less well investigated in University experiments than their external BCI counterparts. A number of animal trials have though been carried out and the more pertinent have been indicated here along with the relevant human trials and practice. In particular the focus of attention has been given to the embodiment of grown neural tissue within a technological body. Whilst only 1,000 or so neurons are involved this presents an interesting research area in a number of ways. But once the number of such neurons used increases 1,000 or 1,000,000-fold, it also raises enormous philosophical and ethical issues. For example is the robot 'thinking' and what rights should it have?

The potential for BCI applications for individuals who are paralysed is enormous, where cerebral functioning to generate command signals is functional despite the motor neural pathways being in some way impaired – such as in Lou Gehrig's disease. The major role is then either one of relaying a signal of intention to the appropriate actuator muscles or to reinterpret the neural signals to operate technology thereby acting as an enabler. In these situations no other medical 'cure' is available, something which presents a huge driver for an invasive implant solution for the millions of individuals who are so affected. Clearly though, bidirectional signalling is important, not only to monitor and enact an individual's intent but also to provide feedback on that individual's resultant interaction with the real world. For grasping, walking and even as a defensive safety stimulant, feedback is vital. This paper has therefore focussed on such studies.

Where invasive interfaces are employed in human trails, a purely therapeutic scenario often exists. In a small number of instances, such as use of the microelectrode array as an interface, an individual has been given different abilities, something which opens up the possibilities of human enhancement. These latter cases however raise more topical ethical questions with regard to the need and use of a BCI. What might be seen as a new means of communication for an individual with an extreme form of paralysis or a new sensory input for someone who is blind, opening up a new world for them, can also be seen as an unnecessary extra for another individual, even though it may provide novel commercial opportunities. What is therapy for one person may be regarded as an enhancement or upgrading for another.

Whilst there are still many technical problems to be overcome in the development of BCIs, significant recent experimental results have indicated that a sufficient technological infrastructure now exists for further major advances to be made. Although a more detailed understanding of the underlying neural processes will be needed in the years ahead, it is not felt that this will present a major hold up over the next few years, rather it will provide an avenue of research in which many new results will shortly appear through trials and experimentation, possibly initially through animal studies although it must be recognised that it is only through human studies that a full analysis can be made and all encompassing conclusions can be drawn. Nevertheless the topic opens up various ethical questions that need to be addressed and as such, research in this area should, I believe, only proceed in light of a pervasive ethical consensus.

ACKNOWLEDGEMENTS

The Author would like to acknowledge the considerable assistance and input of the Consultant Neurosurgeons Mr. Peter Teddy, Mr. Amjad Shad, Mr. Ali Jamous and Mr. Tipu Aziz and researchers Iain Goodhew, Mark Gasson, Ben Whalley and Ben Hutt. Ethical approval for the author's research was obtained from the Ethics and Research Committee at the University of Reading, UK and with regard to the neurosurgery aspect, the Oxfordshire National Health Trust Board overseeing the Radcliffe Infirmary, Oxford, UK.

REFERENCES

Branner, A. and Normann, R. (2000) A multielectrode array for intrafascicular recording and stimulation in the sciatic nerve of a cat, *Brain Research Bulletin*, Vol.51, pp.293-306.

Chapin, J.K. (2004) Using multi-neuron population recordings for neural prosthetics. *Nature Neuroscience*, 7, 452-454.

Carmena, J., Lebedev, M., Crist, R., O'Doherty, J., Santucci, D., Dimitrov, D. Patil, P., Henriquez, C. and Nicolelis, M. (2003) Learning to control a brain-

machine interface for reaching and grasping by primates, *Plos Biology*, Vol.1, Issue.2, article number e2.
Dobelle, W. (2000) Artificial vision for the blind by connecting a television camera to the visual cortex, ASAIO J, Vol.46, pp.3-9.
Donoghue, J. (2002) Connecting cortex to machines: recent advances in brain interfaces, *Nature Neuroscience Supplement*, Vol.5, pp.1085-1088.
Donoghue, J., Nurmikko, A., Friehs, G. And Black, M. (2004) Development of a neuromotor prosthesis for humans, Chapter 63 in *Advances in Clinical Neurophysiology, Supplements to Clinical Neurophysiology*, Vol.57, pp.588-602.
Finn, W. and LoPresti, P. (eds.) (2003) Handbook of Neuroprosthetic methods, CRC Press.
Friehs, G., Zerris, V., Ojakangas, C., Fellows, M. and Donoghue, J. (2004) Brain-machine and brain-computer interfaces, *Stroke*, Vol.35, Issue.11, pp.2702-2705.
Gasson, M., Hutt, B., Goodhew, I., Kyberd, P. and Warwick, K. (2005) Invasive neural prosthesis for neural signal detection and nerve stimulation, Proc. International Journal of Adaptive Control and Signal Processing, Vol.19, No.5, pp.365-375.
Gasson, M., Wang, S., Aziz, T., Stein, J. and Warwick, K. (2005) Towards a demand driven deep brain stimulator for the treatment of movement disorders, Proc. 3rd IEE International Seminar on Medical Applications of Signal Processing, pp.16/1-16/4.
Grill, W. and Kirsch, R. (2000) Neuroprosthetic applications of electrical stimulation, *Assistive Technology,* Vol.12, Issue.1, pp.6-16.
Hinterberger, T., Veit, R., Wilhelm, B., Weiscopf, N., Vatine, J. and Birbaumer, N. (2005) Neuronal mechanisms underlying control of a brain-computer interface, *European Journal of Neuroscience*, Vol.21, Issue.11, pp.3169-3181.
Kennedy, P., Bakay, R., Moore, M., Adams, K. and Goldwaith, J. (2000) Direct control of a computer from the human central nervous system, *IEEE Transactions on Rehabilitation Engineering*, Vol.8, pp.198-202.
Kennedy, P., Andreasen, D., Ehirim, P., King, B., Kirby, T., Mao, H. and Moore, M. (2004) Using human extra-cortical local field potentials to control a switch, *Journal of Neural Engineering*, Vol.1, Issue.2, pp.72-77.
Mann, S. (1997) Wearable Computing: A first step towards personal imaging, *Computer*, Vol. 30, Issue.2, pp. 25-32.
Nicolelis, M., Dimitrov, D., Carmena, J., Crist, R., Lehew, G., Kralik, J. and Wise, S. (2003) Chronic, multisite, multielectrode recordings in macaque monkeys, *Proc. National Academy of the USA*, Vol.100, Issue.19, pp.11041-11046.
Penny, W., Roberts, S., Curran, E., and Stokes, M. (2000) EEG-based communication: A pattern recognition approach, *IEEE Transactions on Rehabilitation Engineering*., Vol. 8, Issue.2, pp. 214-215.
Pinter, M., Murg, M., Alesch, F., Freundl, B., Helscher, R. And Binder, H. (1999) Does deep brain stimulation of the nucleus ventralis intermedius affect postural control and locomotion in Parkinson's disease?, *Movement Disorders*, Vol.14, Issue.6, pp.958-963.
Reger, B., Fleming, K., Sanguineti, V., Simon Alford, S., Mussa-Ivaldi, F. (2000) Connecting Brains to Robots: an artificial body for studying computational properties of neural tissues, *Artificial life*, Vol.6, Issue.4, pp.307-324.
Rizzo, J., Wyatt, J., Humayun, M., DeJuan, E., Liu, W., Chow, A., Eckmiller, R., Zrenner, E., Yagi, T. and Abrams, G. (2001) Retinal Prosthesis: An encouraging first decade with major challenges ahead, *Opthalmology,* Vol.108, No.1.
Roitberg, B. (2005) Noninvasive brain-computer interface, *Surgical Neurology*, Vol.63, Issue.3, p.195.
Warwick, K. (2004) I Cyborg, University of Illinois Press.
Warwick, K., Gasson, M., Hutt, B., Goodhew, I., Kyberd, P., Andrews, B., Teddy, P., Shad, A. (2003) The application of implant technology for cybernetic systems. *Archives of Neurology*, 60 (10), pp.1369-1373.
Warwick, K., Gasson, M., Hutt, B., Goodhew, I., Kyberd, P., Schulzrinne, H. and Wu, X. (2004) Thought Communication and Control: A First Step Using Radiotelegraphy, *IEE Proceedings on Communications,* Vol.151, No. 3, pp 185-189.
Warwick, K., Gasson, M., Hutt, B. and Goodhew, I. (2005) An Attempt to Extend Human Sensory Capabilities by means of Implant Technology. Proc. IEEE Int. Conference on Systems, Man and Cybernetics, Hawaii.
Wolpaw, J., McFarland, D., Neat, G. and Forheris, C. (1990) An EEG based brain-computer interface for cursor control, *Electroencephalogr. Clin. Neurophysiol*., Vol.78, pp.252-259.
Pan, S., Warwick, K., Gasson, M., Burgess, J., Wang, S., Aziz, T. and Stein, J. (2007) Prediction of parkinson's disease tremor onset with artificial neural networks", Proc. IASTED Conference BioMed 2007, Innsbruck, Austria, pp.341-345.
Warwick, K. (2007) The promise and threat of modern cybernetics, Southern Medical Journal, Vol.100, Issue.1, pp.112-115.
Warwick, K. and Gasson, M.N. (2004) Practical Interface Experiments with Implant Technology, in "Computer Vision in Human-Computer Interaction", Sebe, N., Lew, M.S and Huang, T.S. (eds.), Lecture Notes in Computer Science, Vol. 3058, pp.7-16.Xie, S., Yang, Z. and Yang, Y. (2004) Brain-computer interface based on event-related potentials during imitated natural reading, International Journal of Psychology, Vol.39, Issue.5-6, Supplement S., p.138.
Yoo, S., Fairneny, T., Chen, N., Choo, S., Panych, L., Park, H., Lee, S. and Jolesz, F. (2004) Brain-computer interface using fMRI: spatial navigation by thoughts, *Neuroreport*, Vol.15, Issue.10, pp.1591-1595.
Yu, N., Chen, J. and Ju, M. (2001) Closed-Loop Control of Quadriceps/Hamstring activation for FES-Induced Standing-Up Movement of Paraplegics, *Journal of Musculoskeletal Research*, Vol. 5, No.3, pp.173-184.

ANALYSIS AND MODELS OF BRAIN EPILEPTIC ACTIVITIES

Fernando Henrique Lopes da Silva
University of Amsterdam, The Netherlands

Abstract: The essence of epilepsy is the sudden occurrence of a qualitative change in the behaviour of neuronal networks of some specific areas of the brain. In general we may assume that neuronal networks possess multistable dynamics. We may simplify this concept considering the case that a neuronal network may display, at least, two dynamical states: an interictal state characterised by a normal on-going neural activity, as revealed in the Eletcroencephalogram of Magnetoencephalogram (EEG, MEG), that may be apparently random, and another one – the ictal state - that is characterised by the sudden occurrence of synchronous oscillations, most commonly with large amplitude. The latter becomes manifest as a paroxysmal change of behaviour and /of the state of consciousness of a patient, i.e. an epileptic seizure. In the terminology of the mathematics of non-linear systems, we may state that a neuronal network behaves as a bistable system with two attractors, to which the system converges depending on initial conditions and on the system's parameters.

We propose schematically that the transition between the normal on-going to the seizure activity can take place according to three basic models: Model I – a transition may occur due to random fluctuations of some system's parameters. These transitions are thus unpredictable. Models II and III – a transition may result from a gradual change of some unstable parameters, either due to endogenous (model II) or exogenous (model III). In these cases the change of parameter values causes a deformation of the attractor resulting in a transition from the basin of the attractor corresponding to the normal state, to the attractor corresponding to the seizure dynamical state. Some experimental findings obtained in different cases of epilepsy, both in human and in animals, are compatible with each of these 3 models. Some examples of these cases are illustrated.

BRIEF BIOGRAPHY

Fernando Henrique Lopes da Silva received his Medical Degree from the University of Lisbon in 1959, got his Ph.D. from the University of Utrecht in 1970, and in 1980 was appointed Full Professor of General Physiology at the Faculty of Science at the University of Amsterdam (since 2002 part of the Swammerdam Institute for Life Sciences). From 1993 to 2000 he was Director of the newly created Institute of Neurobiology of the University of Amsterdam, and member of the Scientific Directorate of the Graduate School Neurosciences Amsterdam. In 2000, when he reached the retirement age of 65, he became Emeritus Professor of the same University, and has at present a free-lance contract with the Swammerdam Institute for Life Sciences.

Since 1970, he supervised a large number of student trainees from different Universities and Faculties: Medical, Biology, Sciences, (Bio-medical) Engineering. Supervised 65 Ph.D. students (up to December 2006).

His research interests are centred on the biophysical aspects of electrical activity of the brain and the functional organization of neuronal networks, namely of the cerebral cortex and the limbic system, with a special interest in the generation and functional significance of brain rhythmic activities. He published more than 220 papers in peer-reviewed journals and contributed Chapters to 10 multi-authored books (of 6 he is co-editor), among which the Handbook "Electroencephalography: Basic principles, clinical applications and related fields", Niedermeyer, E. and Lopes da Silva, F.H. (Eds), published by Lippincott, Williams and Wilkins, Baltimore; 5 Editions: 1982, 1987, 1993, 1998, 2004. In addition he contributed chapters to to the Encyclopedia of Neuroscience (George Adelman, Barry H. Smith. Eds), Elsevier Science, 2003 (3rd edition), to the Encyclopedia of the Human Brain (Ed. V. S. Ramachandran), Academic Press, 2002, and to The Handbook of Brain Theory and Neural Networks (Ed. Michael A. Arbib), The MIT Press, 2003 (2nd edition).

Selection of Scientific Awards

- 1975 He received the Winkler Medal from the Netherlands Association for Neurology for scientific contributions in the field of neurosciences.
- 1985 Elected member of the Royal Netherlands Academy of Arts and Sciences.
- 1990 "Lord Adrian" Lecturer at the 12th World Congress of Electroencephalography and Clinical Neurophysiology in Rio de Janeiro, Brazil.
- 1992 Honorary President of the VIIth European Congress of Clinical Neurophysiology, Budapest, Hungary.
- 1995 Honorary Life Member of The British Society for Clinical Neurophysiology (Formerly The EEG Society), London, United Kingdom.
- 1997 Doctor Honoris Causa of the University of Lisbon (Portugal).
- 1997 Special "Berger" Lecturer at the 14th International Congress of EEG and Clinical Neurophysiology in Florence, Italy.
- 1999 Recipient of the Herbert H. Jasper Award, selected by the American Clinical Neurophysiology Society for his "lifetime of outstanding contributions to the field of clinical neurophysiology.''
- 2000 Recipient of the 'Storm van Leeuwen/ Magnus Prize' of the Dutch Society of Clinical Neurophysiology.
- 2000 Honorary member of the Portuguese Society of Electroencephalography and Clinical Neurophysiology.
- 2002 Recipient of the Ragnar Granit Prize for his work on the field of Bioelectromagnetism.
- 2002 Doctor Honoris Causa of the University of Porto (Portugal).
- 2004 Recipient of the first Prize "Universidade de Coimbra" for a (sic) "person of Portuguese nationality who has made a particular relevant and innovative contribution in the fields of culture or science."

General Honors

- 2000 High Officer of the Order of Santiago da Espada, for outstanding achievements in the field of Science/Art/Literature, awarded by the President of the Republic of Portugal.
- 2001 Knight of the Order of the 'Nederlandse Leeuw' awarded by the Queen of the Netherlands in appreciation for his achievements in science.

FROM THE BENCH TO THE BEDSIDE

The Role of Semantics in Enabling the Vision of Translational Medicine

Vipul Kashyap
Partners HealthCare System, Clinical Informatics R&D, USA

Abstract: Biomedical research and healthcare clinical transactions are generating huge volumes of data and information. At the same time, the results of biomedical research in the form of new molecular diagnostic tests and therapies are being increasingly used in the context of clinical practice. There is a critical need to speed "translation" of genomic research insights into clinical research and practice. In this talk, we will discuss challenges faced by a healthcare enterprise in realizing the vision of Translational Medicine, such as:

- The need to create structured and semantic representations of genotypic and phenotypic data such as clinical observations and molecular diagnostic tests.
- The need for cost-effective and incremental data integration for combining genotypic and phenotypic information at the point of care.
- The need for actionable decision support for suggesting molecular diagnostic tests and therapies in the context of clinical care.
- The need for knowledge update, propagation and consistency to keep abreast of the rapid pace of knowledge discovery being witnessed in the life sciences, a crucial pre-requisite to reduce the cost of knowledge acquisition and maintenance.

Semantics-based approaches to address the above-mentioned challenges, including the applicability of semantic web standard (RDF, OWL, Rules); and issues related to the value proposition of these technologies will be presented.

BRIEF BIOGRAPHY

Vipul Kashyap, PhD is a Senior Medical Informatician in the Clinical Informatics Research & Development group at Partners HealthCare System and is currently the chief architect of a Knowledge Management Platform that enables browsing, retrieval, aggregation, analysis and management of clinical knowledge across the Partners Healthcare System. Vipul received his PhD from the Department of Computer Science at Rutgers University in New Brunswick in the area of metadata and semantics-based knowledge and information management. He is also interested in characterization of the value proposition of semantic technologies in the enterprise context. Before coming to Partners, Vipul has held positions at MCC, Telcordia (Bellcore) and was a fellow at the National Library of Medicine. Vipul has published 2 books on the topic of Semantics, 40-50 articles in prestigious conferences and journals; and has participated in panels and presented tutorials on the topic of semantic technologies. Vipul sits on the technical advisory board of an early stage company developing semantics-based products, and represents Partners on the W3C advisory committee and the HealthCare Information Technology Standards Panel (HITSP).

THE CANCER INFORMATICS ECOSYSTEM

A Case Study in the Accretion of Federated Systems based on Service Oriented Architectures, Semantic Integration and Computing Grids

David Hall
Research Triangle Institute in North Carolina, USA

Abstract: Information technology is playing an increasingly critical role in health and life sciences research due to the profound expansion in the scope of research projects in the post-genomic age. Robust data management and analysis systems are becoming essential enablers of these studies. Driven by funding agency requirements, funding opportunities, and grass roots organizing, efforts are underway to develop standards and technologies to promote large-scale integration of publicly-funded systems and databases including infrastructure developed for individual studies. Predicted benefits include an enhanced ability to conduct meta-analyses, an increase in the usable lifespan of data, a funding agency-wide reduction in the total cost of IT infrastructure, and an increased opportunity for the development of third party software tools. This presentation will critically examine efforts towards developing publicly-accessible interoperable and distributed production systems in the health and life sciences via ontologies, formal metadata, service oriented architectures, and grid computing models with a focus on several projects under the direction of the author in the area of cancer informatics.

BRIEF BIOGRAPHY

David Hall is a Senior Software Project Leader at RTI International based in North Carolina, USA. He leads teams of up to 30 developers implementing computer systems that support large biomedical and biotechnological research enterprises in cancer research, drug discovery, genetic epidemiology, and plant biotechnology. His area of interest is the practical application of bioinformatics and medical informatics methods, technologies, and standards in the development of production software. Particular topics of interest include data visualization, semantic integration, systems integration, and high performance computing. Recent clients include the US National Institutes of Health, GlaxoSmithKline, Syngenta, and Duke University. Data systems developed by David's group manage clinical and research data for nearly one million patients. Applications include data warehouses, metadata registries, workflow systems, high resolution image databases, analytical applications, and web services. David is currently Principal Investigator of the Informatics Support Center for the National Cancer Institute's Breast and Colon Cancer Family Registries. He holds a Ph.D. in Genetics from the University of Georgia and a B.S. in Computer Science from Wake Forest University.

ICT AND PERSONS WITH DISABILITIES
The Solution or the Problem?

Albert M. Cook
Faculty of Rehabilitation Medicine, University of Alberta, Edmonton Alberta, Canada
al.cook@ualberta.ca

Keywords: Assistive technologies, information and computer technologies, persons with disabilities.

Abstract: In order to lead full and productive lives, persons with disabilities need to have the same access to information and communication systems as the rest of the population. Advances in information and communication technologies (ICT) are occurring quickly, and the capability of technologies to meet the needs of persons with disabilities is growing daily. Future developments in assistive technologies (AT) and the successful application of these technologies to meet the needs of people who have disabilities are dependent on exploitation of these ICT advances. AT also involves the development of specialized interfaces such as the brain computer interface (BCI), adaptive interfaces that accommodate for changes in the user's physical skills, cognitive interfaces that allow understanding of the human technology interface by individuals with intellectual disabilities and systems that accommodate for user needs based on environmental sensing (e.g., GPS interfaces) and downloading of profiles to meet specific user needs. Universal Design (or design for all) calls for the design of products and environments to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design. In the physical world this often means ramps, curb cuts and other adaptations to the built environment to accommodate individuals who have disabilities. In the ICT world the barriers to access are technological, and the goal for ICT universal design is to have an environment with enough embedded intelligence to be easily adaptable to the varying cognitive, physical and sensory skills of a wide range of individual's in order to meet their productivity, leisure and self care needs. If ICT advances are not adaptable enough to be accessible to persons with disabilities it will further increase the disparity between those individuals and the rest of the population leading to further isolation and economic disadvantage. On the other hand, availability of these technologies in a transparent way will contribute to full inclusion of individuals who have disabilities in the mainstream of society.

BRIEF BIOGRAPHY

Dr. Albert Cook is Professor of Speech Pathology and Audiology and currently Dean of the Faculty of Rehabilitation Medicine and Chair of the Health Sciences Council at the University of Alberta. Dr. Cook has worked with interdisciplinary teams to develop assistive devices and to assess the effectiveness of technology being used by persons with disabilities. Dr. Cook is also associated with the I CAN Centre the Glenrose Rehabilitation Hospital. He was formerly Professor of Biomedical Engineering at California State University, Sacramento where he established the graduate program in biomedical engineering and directed it for 12 years. He also served as Co-Director of the Assistive Device Center in Sacramento, California, helping over 500 persons with disabilities to identify and obtain assistive technologies.

He received his Bachelor of Science in Electrical Engineering at the University of Colorado, a Masters in Bioengineering and his doctorate from the University of Wyoming He is a member of Tau Beta Pi, Phi Kappa Phi and Gold Key honorary societies.

Dr. Cook co-authored with Janice Polgar, OTR, Cook and Hussey's Assistive Technologies: Principles and Practice 3nd edition, published in October 2007 by Elsevier. He has co-edited three other textbooks with John Webster and others and has written numerous chapters in rehabilitation and biomedical engineering texts and monographs.

Dr. Cook's research interests include augmentative and alternative communication, biomedical instrumentation and assistive technology design, development and evaluation. His most recent

research has focussed on the use of robotics with young children who have severe disabilities to develop and assess cognitive and linguistic skills. He has US and foreign patents and numerous publications and conference presentations in these areas. He has been principal investigator on research and training grants in augmentative communication, assistive technologies and biomedical engineering.
Dr. Cook is Past-President and Fellow of RESNA, a major professional society for assistive technology practitioners in North America. He has also served in national united States positions in the Institute of Electrical and Electronic Engineers Engineering in Medicine and Biology Society, the American Society for Engineering Education, the Biomedical Engineering Society, the International Society for Augmentative and Alternative Communication and the Association for the Advancement of Medical Instrumentation. Dr. Cook is a registered professional engineer (electrical) in California.

1 ICT AND PERSONS WITH DISABILITIES TECHNOLOGY AND PROGRESS

Societal Progress requires change much of which is accomplished through advances in technology. In his book, *A Short History of Progress*, Ronald Wright *(*2004) points out that this characteristic has been true for millions of years as societies have advanced through greater utilization of technology.

Wright goes on to describe the problems that technology typically creates such as over consumption, environmental ruin, and separation of classes. These problems are amplified for people who have disabilities, and they lead to a gap in the access to work, self care and community participation for persons with disabilities compared to the general population. Since people with disabilities often depend on technologies for societal participation, the lack of availability of accessible technology or the obsolescence of accessible technologies isolates them further. This is an extension of the concept of the "digital divide" that separates people along socioeconomic lines based on their access to ICT. I refer to it as the "disability gap".

2 ADVANCES IN INFORMATION AND COMMUNICATION TECHNOLOGIES (ICT)

The 21st Century is characterized by a continuous move from a machine-based to a knowledge based economy (Ungson & Trudel, 1999). In this shift, the basis of competence is changing to knowledge skills from machine skills. Information currently amounts to 75% of value added to products This will continually increase, and connectivity will be the key to business success. There is also a move from a regional or national scope of business influence to a global scope, in which virtual networks dictate organizational structures.

Key players in business development are becoming communication suppliers with the move from host-based to network based systems. Telephone, cable TV and internet service providers control commercial growth. Along with these changes networks will become more graphically-based moving increasingly from text-based systems. In order to lead full and productive lives, persons with disabilities need to have the same access to this new information and communication system as the rest of the population.

2.1 What Can we Expect from Technology in the Next 20 Years?

The cost of information technology is continually dropping for comparable or increased computing power and speed. There is also a greater understanding of the biological/physical interface for the control of computers. The human computer interface (HCI) is being developed to be more human-like, more user oriented and more intelligent-providing additional capabilities for searching, processing and evaluating information.

There are a number of changes that are likely to occur over the next few years (Applewhite, 2004). There will be an increase in automated transactions between individuals and organizations enabling people to complete all transactions without face-toface interactions. It is expected that we will achieve equalized access to the web and information between the developed and developing world. Embedded systems will dramatically increase with application such as "intelligence in the doorknob" that recognizes the owner and doesn't require key manipulation. We are likely to see much greater

understanding of the biological to physical interface for the control of computers.

2.2 Changes in Mainstream Tech with AT Implications

There are many examples of emerging mainstream technologies with potential for assisting people with disabilities to access ICT systems. A few of these are described in this section.

Display-based assistive technologies present an array of choices for a user to select from (Cook & Polgar, 2007). This often referred to as scanning since the choices are highlighted sequentially and then chosen using some sort of gross movement. One of the problems associated with this approach is that there must be a physical display for making selections. This often requires the overall system to be larger and more bulky or places a display between a user and a communication partner. A new development is a direct retinal display that creates image that overlays view of real object (Lewis, 2004). The retinal display is low powered because it is shined on retina directly. Scanning light into the eye allows the image to overlay an object such as a communication partner's face-enabling eye contact and small size. The scanning array could be the retinal image, since display scans across the retina power levels can be kept low kept low for safety.

Another development is 3-D displays that create a more intuitive view of objects, events and activities (Lewis, 2004). This type of display may be helpful to individuals who have cognitive disabilities. It might also create new challenges for individuals with visual limitations.

Embedded automatic speech recognition is being developed for PDAs because of the need for keyboards with more and more functions and the limitations of very small keyboards (Kumagai, 2004). This feature could be very useful to reduce individuals who have limited hand function or for those who cannot see the keyboard to make entries.

3 MEETING THE ICT NEEDS OF PERSONS WITH DISABILITIES

Over the centuries, our ability to make tools is what distinguishes us as human, but our tools ultimately control us by making us dependent on them (Wright, 2004). This dependence is less optional for people who have disabilities

3.1 Impact of Technology Advances on People who have Disabilities

Technology advances increase the gap between people who have disabilities and those who don't (Wright, 2004). All societies become hierarchical with an upward concentration of wealth (including aggregations of technology tools) that ensures that "there can never be enough to go around", and this disparity contributes to the "digital divide" and the "disability gap". As advances occur more quickly, the gap widens faster and the people who are poor and/or disabled loose out even more completely and faster. This is a characteristic of cultural and societal "progress" over centuries-technology drives change, and creates both positive and negative outcomes in the process

The prognosis is not good for people with disabilities unless there is considerable effort to keep them connected to ICT and thereby to commerce, employment and personal achievement. There two fundamental approaches to this problem (1) make mainstream technologies accessible to people who have disabilities, or (2) design special purpose technologies specifically for people with disabilities. The former approach is referred to as *universal design* or *design for all*. The second approach utilizes *assistive technologies*.

3.2 Implications for Assistive Technologies

Access to ICT for people with disabilities is a significant global problem, and it has major implications for assistive technologies. There is a constant challenge to keep ICT systems accessible to persons who have disabilities as mainstream advances occur and adaptations become potentially incompatible with the new systems. Communication technologies change rapidly, and each change may result in the need to re-design accessible interfaces. We are closer to goal of having assistive technology adaptations available when the mainstream consumer product ships, but there are still many problems with "workarounds" necessary to make mainstream operating system, productivity software and internet access accessible to people with disabilities.

Development and maintenance of access to ICT must be driven by the needs of people with disabilities. Developments which broaden the scope, applicability and usability of the human technology

interface will be driven, at least in part by the needs of people who have disabilities.

The Internet (e-mail and chat rooms) have the advantage of anonymity, and this can be a major benefit to individuals who have disabilities. Because the person's disability is not immediately visible, people who have disabilities report that they enjoy establishing relationships with people who experience them first as a person and then learn of their disability. For example, Blackstone, (1996) describes some of the advantages of e-mail for individuals who have disabilities. Since the receiver of the message reads it at a later time composition can be at a slower speed. The person with a disability can communicate with another person without someone else being present, establishing a greater sense of privacy than situations in which an attendant is required. It is also possible to work form any location-avoiding some transportation problems

3.3 Universal Design

Increasingly, commercial products are being designed to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design (NC State University, The Center for Universal Design, 1997).

3.3.1 General Principles of Universal Design

Features are built into products to make them more useful to persons who have disabilities (e.g., larger knobs; a variety of display options--visual, tactile, auditory; alternatives to reading text--icons, pictures) are built into the product. This is much less expensive than modifying a product after production to meet the needs of a person with a disability. The North Carolina State University Center for Universal Design, in conjunction with advocates of universal design, have compiled a set of principles of universal design, shown in Box 1. This center also maintains a Web site on universal design (www.design.ncsu.edu/cud).

3.3.2 Universal Design for ICT

In universal design for ICT the barriers are technological rather than political and economic barriers that characterize architectural and commercial product design (Emiliani, 2006). The goal of universal design for ICT is to have an environment with enough embedded intelligence to be easily adaptable. The features of future information services are that there will be no clearly predefined service and little distinction between interpersonal communication and access to information. Services will need to be highly interactive, inherently multimedia, sensory multimodal (i.e., access via auditory or visual means is equally possible). To achieve this cooperation between users or representatives of users is critical in a variety of contexts of use. The overall goal is to have access to information involving communities of users with a wide range of motor, sensory and cognitive skills.

ONE: EQUITABLE USE
The design is useful and marketable to people with diverse abilities.
TWO: FLEXIBILITY IN USE
The design accommodates a wide range of individual preferences and abilities.
THREE: SIMPLE AND INTUITIVE USE
Use of the design is easy to understand, regardless of the user's experience, knowledge, language skills, or current concentration level.
FOUR: PERCEPTIBLE INFORMATION
The design communicates necessary information effectively to the user, regardless of ambient conditions or the user's sensory abilities.
FIVE: TOLERANCE FOR ERROR
The design minimizes hazards and the adverse consequences of accidental or unintended actions.
SIX: LOW PHYSICAL EFFORT
The design can be used efficiently and comfortably and with a minimum of fatigue.
SEVEN: SIZE AND SPACE FOR APPROACH AND USE
Appropriate size and space is provided for approach, reach, manipulation, and use regardless of user's body size, posture, or mobility.

Box 1: Principles of Universal Design From North Carolina State University, The Center for. Universal Design, 1997.

In addition to Universal Design for ICT, access to capabilities of mainstream technologies includes individualized assistive technologies that are easily – customized. This in return requires an increased understanding of the biological/physical interface for the control of assistive technologies and expanded availability of embedded systems networks.

3.4 A Working Definition of Assistive Technologies

The *International Classification of Functioning, Disability and* Health (ICF) is a system developed by the World Health Organization (WHO) that is designed to describe and classify health and health related states. These two domains are described by

body factors (body structures and functions) and individual and societal elements (activities and participation) (WHO, 2001). The ICF recognizes two contextual factors that modify health and health related states: the environment and personal factors (WHO, 2001). Environmental elements include assistive technologies in relation to activities of daily living, mobility, communication, religion and spirituality as well as in specific contexts such as education, employment and culture, recreation and sport (WHO, 2001). Other environmental elements such as access to public and private buildings, and the natural and built outdoor environments, also have implications for assistive technologies.

A commonly used definition of assistive technology is from the Technical Assistance to the States Act in the United States (Public Law (PL) 100-407): ***Any item, piece of equipment or product system whether acquired commercially off the shelf, modified, or customized that is used to increase, maintain or improve functional capabilities of individuals with disabilities.***

3.4.1 Hard and Soft Technologies

Odor (1984) has distinguished between hard technologies and soft technologies. Hard technologies are readily available components that can be purchased and assembled into assistive technology systems. The main distinguishing feature of hard technologies is that they are tangible. On the other hand, soft technologies are the human areas of decision making, strategies, training, concept formation, and service delivery as described earlier in this chapter. Soft technologies are generally captured in one of three forms: (1) people, (2) written, and (3) computer (Bailey, 1997). These aspects of technology, without which the hard technology cannot be successful, are much harder to obtain. Soft technologies are difficult to acquire because they are highly dependent on human knowledge rather than tangible objects. This knowledge is obtained slowly through formal training, experience, and textbooks such as this one. The development of effective strategies of use also has a major effect on assistive technology system success. Initially the formulation of these strategies may rely heavily on the knowledge, experience, and ingenuity of the assistive technology practitioner. With growing experience, the assistive technology user originates strategies that facilitate successful device use. There is a false belief that progress is solely driven by "hard" technological change The gap between the general public and persons with disabilities can only be closed by gains in both soft and hard technologies

3.4.2 Mainstream Technologies to Specially Designed Technologies: A Range of Options

As illustrated in Figure 1, the needs of people with disabilities can be met in a number of ways. Off the shelf "standard" (i.e., mainstream technologies) commercially available devices (especially those designed using the principles of universal design) can often be used by people with a variety of disabilities. For example, standard personal computers designed for the general population are often used by persons with disabilities. Sometimes these need to be modified however, to make them useable. Another type of commercially available device is one that is mass-produced but specifically designed for individuals with disabilities *(special commercially available devices)*. These devices often need to be modified to meet the needs of a specific individual. Our goal is to reduce the amount of modification necessary and to make mainstream technologies as accessible as possible. However, there will always be a portion of the disabled population that will require specifically designed assistive technologies.

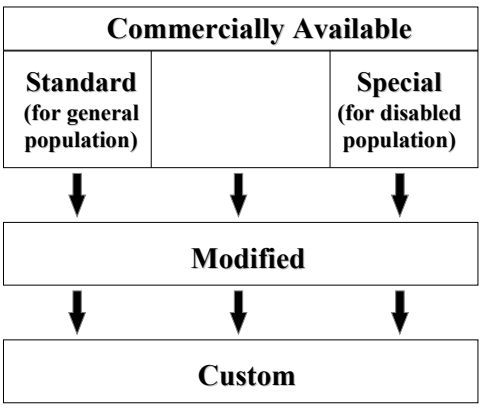


Figure 1: This diagram shows the progression from commercially available devices for the general population and commercially available devices for special populations to modified devices and custom devices. From Cook and Polgar, (2007).

3.5 The Human Technology Interface for ICT

3.5.1 General Concepts

It is estimated that as many as 40 million persons in the United States alone have physical, cognitive, or sensory disabilities (Lazzaro, 1999). The world-wide impact is significantly larger. If these people are to compete on an equal basis with non-disabled individuals, then it is extremely important that the internet be accessible to all. As the internet becomes more and more dependent on multimedia representations involving complex graphics, animation, and audible sources of information, the challenges for people who have disabilities increase. In order for access to the Internet to be useful to people with disabilities, the accessibility approach must be independent of individual devices. This means that users must be able to interact with a *user agent* (and the document it renders) using the input and output devices of their choice based on their specific needs. A **user agent** is defined as software to access Web content (www.w3.org/wai). This includes desktop graphical browsers, text and voice browsers, mobile phones, multimedia players, and software assistive technologies (e.g., screen readers, magnifiers) that are used with browsers. The person with a disability interacts with technology through the Human Technology Interface (HTI) (Cook and Polgar, 2007).

The graphical user interface (GUI) has both positive and negative implications for persons with disabilities. The positive features are those that apply to non-disabled users (e.g., use of icons, recognition rather than recall memory, screen icons for the same task look the same, operations such as opening and closing files are always the same). The GUI is the standard user interface because of its ease of operation for novices and its consistency of operation for experts. The latter ensures that every application behaves in basically the same way. People with motor disabilities may not have the necessary physical (eye-hand coordination) and visual skills to navigate the GUI. Modification of the GUI to allow specialized access (see Figure 1) can also be more challenging for GUI-based operating systems.

As networks are expanded and more devices (e.g., cell phones, PDAs) have open architectures, it will be possible to download profiles, adaptations and special instructions that enable adaptable systems to be developed to meet the needs of people who have disabilities. Some examples are (1) trainable hearing aids that adjust automatically to the environments in which they are used; (2) a "Smart House" that assesses occupants current state and the state of various home utilities to aid with common activities of daily living, provides feedback should residents become disoriented or confused and report medical emergencies automatically; an orientation and direction finding device that senses the current location (via GPS) and gives directions to a desired location for individuals who cannot read maps because of visual or cognitive disabilities.

3.5.2 Access for Motor Impairment

There are a significant number of people who cannot effectively use standard keyboards, mouse controls or switches. It is likely that we will see a much greater understanding of the biological/physical interface for the control of computers in the future (Applewhite, 2004).

One approach that may offer promise is the brain computer interface (BCI). BCI systems may be grouped into a set of functional components including the input device, amplification, feature extraction, feature translation and user feedback (Mason and Birch, 2003). Signals are mathematically analyzed to extract features useful for control (Fabiani, Mcfarland, Walla, and Pfurtscheller 2004). Features or signals that have been used include slow cortical potentials, P300 evoked potential, sensorimotor rhythms recorded from the cortex and neuronal action potentials recorded within the cortex). A typical task for a user is to visualize different movements or sensations or images.

Another approach to cursor control is the use of a digital camera and image recognition software to track a particular body feature to control an on-screen mouse cursor (Betke, Gips and Fleming, 2002). The most easily tracked feature is the tip of the nose, but the eye (gross eye position not point of gaze), lip, chin and thumb have also been used. Non-disabled subjects used this approach and fund that the camera mouse was accurate but slower than a typical hand-controlled mouse. Using an on-screen keyboard the camera mouse was half as fast as a regular mouse in a typing task, but the accuracy obtained was equivalent on each system. More and more computers have built-in cameras, so the camera mouse requires only software to capture the body feature image and interpret its movement as mouse commands. This may lead to wider application of this technique.

There are many other approaches that are used to provide access to and control over technologies for people with severe motor disabilities (Cook and Polgar, 2007) \. These range form keyboards of various type, to automatic speech recognition to mouse and mouse emulators systems to single and multiple switches.

3.5.3 Access for Cognitive Impairment

Cognitive disabilities include a wide range of skills and deficiencies. Learning disabilities typically involve significant difficulties in understanding or in using either spoken or written language, and these difficulties may be evident in problems with reading, writing, mathematical manipulation, listening, spelling or speaking (Edyburn, 2005). These limitations make it increasingly difficult to access complicated Web sites that may include flashing pictures, complicated charts, and large amounts of audio and video data. While there are assistive technologies that are specifically designed to address these areas (discussed later in this chapter), many of the technological tools are useful for all students, and are part of instructional technology (Ashton, 2005). Even the so-called assistive technologies have features (e.g., multimedia, synthetic speech output, voice recognition input) that are useful to all learners.

For individuals with acquired cognitive disabilities due to injury (e.g., traumatic brain injury) or disease (e.g., stroke (CVA) or dementia) changing features such as font size, background/ foreground color combinations, contrast, spacing between words, letters and paragraphs and using graphics can all improve access to screen-based information. Another technological concept for these individuals is a cognitive prosthesis, which is a custom-designed computer-based compensatory strategy that directly assists in performing daily activities[1]. It may also include additional technologies such as a cell phone, pager, digital camera or low tech approaches

Persons with intellectual disabilities have difficulties with memory, language use and communication, abstract conceptualization, generalization and problem identification/problem solving. Characteristics of the HTI that are important for these individuals include simplicity of operation, capacity of the technology to support repetition, consistency in presentation, and inclusion of multiple modalities (e.g., speech, sounds and graphical representations) (Wehmeyer, Smith and Davies, 2005).

An example of technology designed for cognitive needs is the Planning and Execution Assistant and Trainer (PEAT). It is a PDA-based personal planning assistant designed to assist individuals with cognitive disorders due to brain injury, stroke, Alzheimer's disease, and similar conditions (Levinson, 1997). PEAT employs artificial intelligence to automatically generate plans and also to revise those plans when unexpected events occur. PEAT uses a combination of manually entered schedules and a library of stored scripts describing activities of daily living (e.g., morning routine or shopping). Scripts can be used for both planning and for execution. Planning involves a simulation of the activity with key decision points presented and prompts (auditory and visual) supplied necessary to aid the individual through the planning process. The plan to be executed can be either the stored script or a modified script based on the simulation. The PEAT artificial intelligence software generates the best strategy to execute the required steps in the plan (LoPresti, Mihailidis, and Kirsch, 2004). PEAT also automatically monitors performance, and corrects schedule problems when necessary.

3.5.4 Access for Auditory Impairment

Since web pages are a mixture of text, graphics, and sound, people who are deaf may be prevented from accessing some information unless alternative methods are available. The primary approach for thee individual is the use of the Microsoft Synchronized Accessible Media Interchange (SAMI), which allows authors of Web pages and multimedia software to add closed captioning for users who are deaf or hard of hearing. This approach is similar to the use of closed captioning for television viewers. The W3C WAI SMIL (www.w3.org/WAI) is designed to facilitate multimedia presentations in which an author can describe the behavior of a multimedia presentation, associate hyperlinks with media objects, and describe the layout of the presentation on a screen

Trainable hearing aids adjust automatically to the environments in which they are used through access to embedded information networks. This allows automatic adaptation to changing noise levels and environments.

[1] Institute for Cognitive Prosthetics, http://www.brain-rehab.com/ definecp.htm

3.5.5 Access for Visual Impairment

The W3C WAI user agent guidelines are based on several principles that are intended to improve the design of both types of user agents. The first is to ensure that the user interface is accessible. This means that the consumer using an adapted input system must have access to the functionality offered by the user agent through its user interface. Second, the user must have access to document content through the provision of control of the style (e.g., colors, fonts, speech rate, and speech volume) and format of a document. A third principle is that the user agent help orient the user to where he is in the document or series of documents. In addition to providing alternative representations of location in a document (e.g., how many links the document contains or the number of the current link), a well-designed navigation system that uses numerical position information allows the user to jump to a specific link. Finally, the guidelines call for the user agent to be designed following system standards and conventions. These are changing rapidly as development tools are improved.

Communication through standard interfaces is particularly important for graphical desktop user agents, which must make information available to assistive technologies. Technologies such as those produced by the W3C include built-in accessibility features that facilitate interoperability. The standards being developed by the W3C WAI provide guidance for the design of user agents that are consistent with these principles. The guidelines are available on the W3C WAI Web page (www.w3.org/wai).

3.5.6 Other ICT Access

Cellular telephones are becoming more powerful with capabilities approaching that of personal computers. This expanded capability will provide significant advantages for people with disabilities, especially those with low vision or blindness. describes Three changes will be particularly valuable to people who have disabilities:: (1) standard cell phones will have sufficient processing power for almost all the requirement of persons with visual impairments, (2) software will be able to be downloaded into these phones easily, (3) wireless connection to a worldwide network will provide a wide range of information and services in a highly mobile way (Fruchterman, 2003). Because many of these features will be built into standard cell phones the cost will be low and reachable by persons with disabilities. A major advance will occur if the cell phone industry moves away from proprietary software to an open source format providing the basis for a greater diversity of software for tasks such as text-to-speech output, voice recognition and optical character recognition in a variety of languages. Many applications for people with disabilities will be able to be downloaded from the internet. With expanded availability of embedded systems, it will be possible for a user to store their customized programs on the network and download them as needed form any remote location.

Downloading a talking book program into a cell phone can provide access to digital libraries for persons who are blind. Outputs in speech or enlarged visual displays can be added as needed by the user. With a built-in camera and network access a blind person could obtain a verbal description of a scene by linking to on-line volunteers who provide descriptions of images. These applications will depend on the increasing application of universal design in information technology products (Tobias, 2003). These applications include ATMs, cell phones, vending machines and other systems that are encountered on a daily basis (Tobias, 2003).

4 INFRASTRUCTURE FOR FUTURE ACCESSIBILITY

The infrastructure for future accessibility consists of: (1) an expanded, smarter and more available "real" and "virtual" internet, (2) Home automation systems that are smarter and have greater interconnectivity, (3) universal design principles that are applied more widely, (4) alternative approaches for accessing information technologies, and (5) special-purpose assistive technologies.

The Infrastructure for future accessibility will depend on several factors. These include: Web-based virtual systems, home automation, universal design for ICT, alternatives for accessing information technologies and special-purpose assistive technologies. In addition there is n on going need for the development of soft technology tools.

If ICT advances are not adaptable enough to be accessible to persons with disabilities it will further increase the disparity between those individuals and the rest of the population leading to further isolation and economic disadvantage. On the other hand, availability of these technologies in a transparent way will contribute to full inclusion of individuals who have disabilities in the mainstream of society.

5 CONCLUSIONS

The move to the information age offers great promise for persons with disabilities. It also holds great threats for persons with disabilities. Constant vigilance is required to insure that information technologies remain accessible and responsive to the needs of persons with disabilities. The future for persons with disabilities will not be driven by advances in technology, but rather by how well we can take advantage of those advances for the accomplishment of the many tasks of living that require technological assistance

6 SUMMARY

Anticipated changes in technologies coupled with the focus on the social aspects of disability, provide a significant opportunity for major advances in the degree to which individuals with disabilities can participate in all aspects of life, including work, school, leisure and self care.

Technological advances will be particularly important as the percentage of the population that is elderly rises. Concepts from universal design will be important in ensuring that this segment of the population remains active and is able to participate in society. This new group of elderly individuals will also be more experienced with computers and other technologies than their predecessors and they may well demand greater performance and adaptability from both assistive technologies and mainstream ICT (e.g., telephones, internet communication).

The percentage of individuals with long-term disabilities who join the over 65 age group will also increase. These individuals will have been long-term users of assistive technologies, and their experience will have major implications for developments to meet future needs.

While much of what I have described is conjecture, it is based on modest extrapolation from the current state of the art. There are some things that we know with a high degree of certainty. We know that computer systems will be faster, have more memory be smaller and be less expensive for the same or greater functionality. We also know that the communication channel bandwidth will continue to increase allowing much more information and much more sophisticated information processing. Finally, it is clear that people with disabilities will continue to assert their right to fully participate in society.

Technological advances also raise questions for people who have disabilities. The most important of these is whether accessibility will keep pace with technological developments. For example, will assistive technologies for input and output be compatible with the user agents and operating systems of tomorrow. A second major question is whether the needs of persons with disabilities will be a driving force in future technological developments. Will people who have disabilities have to adapt to the existing technologies based on characteristics for non-disabled people or will universal design become a greater reality? In the latter case, adaptations will become less important and accessibility will become the rule rather than the exception.

For people who have disabilities, there are significant implications of emerging information processing technologies. If not closely monitored, these could result in less rather than more access to the new information economy for persons with disabilities. Despite the wider use of universal design principles, there will still be a need for effective assistive technology design and application if individuals with disabilities are to realize the full potential of the new information age.

REFERENCES

Applewhite A: (2004). 40 years: the luminaries. *IEEE Spectrum,* 41(11), 37-58.

Ashton T.M. (2005). Students with learning disabilities using assistive technology in the inclusive classroom, pp. 229-238, in Edyburn D., Higgins K.. & Boone R. (eds.): *Handbook of Special Education Technology Research and Practice,* Whitefish Bay, Wisconsin: Knowledge by Design, Inc.

Bailey, R.W: (1996). *Human performance engineerin.,* ed 2, Upper Saddle River, NJ, , Prentice Hall.

Betke M., Gips J. & Fleming P: (2002). The camera mouse: Visual tracking of body features to provide computer access for people with severe disabilities. *IEEE Trans. Neural Systems and Rehabilitation Engineering,* 10(1): 1-10.

Blackstone S. (1996). The Internet: what's the big deal *Augment Commun News* 9(4):1-5.

Cook A.M. & Polgar J.M. (2007). *Cook and Husey's Assistive Technologies: Principles and Practice*, 3rd ed St. Louis: Elsevier.

Edyburn D.L. (2005). Assistive technology and students with mild disabilities: from consideration to outcome measurement, pp 239-270, in Edyburn D, Higgins K. & Boone R. (eds.): *Handbook of Special Education Technology Research and Practice,* Whitefish Bay, Wisconsin: Knowledge by Design, Inc.,

Emiliani, P.L. (2006). Assistive technology (AT) versus Mainstream Technology (MST): The research perspective. *Tech Disab*, 18, 19-29,

Fabiani G.E., Mcfarland D.J., Wolpaw J.R., & Pfurtscheller G (2004). Conversion of EEG activity into cursor movement by a brain-computer interface (BCI). *IEEE Trans. Neural Systems and Rehab Engr*, 12: 331-338.

Fruchterman J.R: (2003) In the palm of your hand: A vision of the future of technology for people with visual impairments. *J Vis Impair Blindness, 97*(10), 585-591.

Kumagai, J. (2004).Talk to the machine. *IEEE Spectrum,* 39(9): 60- 64.

Lazzaro J.L. (1999). Helping the web help the disabled, *IEEE Spectrum* 36(3): 54-59

Levinson R.L. (1997): The planning and execution assistant and trainer. *J. head trauma rehabilitation*, 12(2): 769-775.

Lewis, J.R. (2004). .In the eye of the beholder*, IEEE Spectrum,* 41(5)24- 28.

LoPresti E.F., Mihailidis A. & Kirsch N. (2004): Assistive technology for cognitive rehabilitation: State of the art. *Neuropsychological Rehabilitation,* 14(1): 5-39.

Mason S.G. & Birch G.E: (2003).A general framework for brain-computer interface design. *IEEE Trans. Neural Systems and Rehab Engr*, 11, 70-85.

Odor P. (1984): Hard and soft technology for education and communication for disabled people, *Proc Int Comp Conf*, Perth, Western Australia,

Tobias J: (2003). Information technology and universal design: An agenda for accessible technology. *J Vis Impair Blindness, 97*(10), 592-601.

Ungson G.R. and Trudel J.D. (1999). The emerging knowledge-based economy. *IEEE Spectrum*, 36(5):60-65.

Wehmeyer M.L., Smith S.J., Palmer SB, Davies D.K. & Stock S.E: (2004). Technology use and people with mental retardation. *International Review of Research in Mental Retardation,* 29: 291-337.

World Health Organization (2001): *International classification of functioning disability and health*-ICF, Geneva, World Health Organization

Wright, RA. (2004). *A Short History of Progress*, Toronto: Anansi Publications.

SHORT PAPERS

INTERACTION OF TECHNOLOGICAL AND INSTITUTIONAL CHANGE IN THE DEVELOPMENT OF AN ELECTRONIC COMMERCE SYSTEM IN CHINA'S PHARMACEUTICAL DISTRIBUTION CHAIN

A Transaction Cost Perspective

Kai Reimers
Faculty of Business and Economics, RWTH Aachen University, Johanniterstr. 22-24, 52064 Aachen, Germany
reimers@wi.rwth-aachen.de

Mingzhi Li
School of Economics and Management, Tsinghua University, 100084 Beijing, P.R. China
limzh@sem.tsinghua.edu.cn

Keywords: Electronic commerce, pharmaceutical industry, transaction costs, institutional change.

Abstract: In this paper, we describe the introduction of electronic commerce into the drug distribution industry in China. This case is especially interesting because massive institutional changes coincide with the introduction of a new technology. For these reasons, it becomes possible to study the interaction of technological and institutional change in real time and in their real-life context. We use two alternative transaction cost-theoretic perspectives on the interaction between institutional and technological change as our theoretical framework. The case study suggests that the rationale which motivates introduction of a new technology in the context of institutional change may be different from the rationale which underlies the practices which develop on the basis on the newly introduced systems and institutions.

1 INTRODUCTION

Studying the interaction of institutional and technological change in real time is generally a difficult endeavour because of the slow speed at which institutions typically change. For this reason, scholars have focused on two empirical strategies, namely historical studies with respect to large-scale institutional changes, usually on the national level, and studies of institutional change on the level of individual companies (where institutions are changed more easily and thus more quickly). Prominent representatives of these two approaches are Alfred Chandler and Oliver Williamson respectively. Large-scale institutional changes become visible only after a substantial period of time has elapsed so that only historical analysis is feasible; studies on the level of individual organizations such as firms can only reveal small-scale institutional changes. While both approaches have yielded impressive results, studying the interactions between institutional and technological changes on an intermediate level in real time would be desirable for the following reason. While institutions exist on multiple levels so that institutional change can also be studied on multiple levels, new technological systems are generally not developed on both the national scale and the level of individual firms but on intermediate levels. This is because development of new technological systems typically requires the cooperation among several organizations, not only because resources and capabilities of individual organizations are too limited but also because application of new systems, at least in the initial stages, requires close cooperation of users and suppliers who are thus actively involved in the design and development process (von Hippel, 1978).

However, institutional changes occurring on the intermediate level are seldom observable in real time because of the slow speed of institutional change on this level. One exception to this rule can be found in fast developing countries which actively experiment with medium-scale institutional changes. China may be the most interesting and relevant context in this regard and within China, the healthcare sector is the

one which currently receives the most attention by policy makers due to its economic as well as social and political importance.

Theoretically, the interaction between technical and institutional change has been addressed from a number of perspectives, including transaction cost theory. However, as North and Wallis (1994) have shown, existing institutional approaches -- including those by Chandler and Williamson -- only consider institutional change as a passive response to technical change. North and Wallis trace this traditional limitation back to an assumption that transaction costs cannot be observed or measured which has prevented scholars from appreciating an important mechanism through which technical change can also contribute to proactive institutional change. According to North and Wallis, in this tradition it is generally assumed that new technology requires institutional adaptation in order to be implemented which, however, will increase unit transaction costs even though the transaction cost minimizing institutional arrangement will be chosen. In contrast, if one accepts that transaction costs can be empirically observed, one will be naturally led to look for those parts of the economy which directly express the level of transaction costs existing in an economy at any given moment in time. North and Wallis have identified this sector as the 'transaction sector' and used it to measure the existing level of transaction costs in an economy (Wallis and North, 1986). By this approach they can show how institutions can be changed proactively as a result of a lowering of unit transaction costs through technical change. Thus, North and Wallis demonstrate two ways in which technical change can lead to institutional change: (1) in order to accommodate the requirements of a new technology, new institutions may have to be created which, however, increase unit transaction costs in an economy; since transformation costs are lowered simultaneously (through technical change), the overall effect can still be beneficial; (2) technical change can lower unit transaction costs and thus enable the creation of new institutional forms.

In this paper we want to explore this dual relationship using the case of a massive institutional reform of the distribution of drugs in China which coincided with the rise of e-commerce and led to a new way of selling and distributing drugs in which both elements of institutional and technological change are intertwined. We want to investigate to which extent emergence of this new structure follows the Williamson-Chandler pattern -- new technology needs new institutional forms which result in higher unit-transaction costs -- and to which extent it can be explained as a result of transaction-cost reducing technological change according to North's framework.

2 THEORETICAL FRAMEWORK

Transaction cost economics takes the transaction as its unit of analysis and therefore operates on an analytical level suitable for studying technological change affecting interaction among economic agents. Williamson (1993) has used a three-layer model to locate and motivate the analytic strategy of transaction cost economics. The focal layer concerns the governance of transactional relationships which is influenced by factors residing on a macro and a micro layer. On the macro layer, institutions such as contract law, property rights and norms and customs affect the comparative costs of governance. Changes on this layer may therefore lead to adjustments of governance structures as some forms of governance become relatively more cost advantageous while others become more costly. On the micro-level, behavioural traits of individuals influence governance forms. Prominent among them in Williamson's analysis are opportunism and bounded rationality. For example, governance forms need to prevent economic agents from exploiting one another should the possibility for such behaviour arise due to unanticipated changes in the economic environment.

While Williamson has not specifically developed his transaction cost-theoretic framework for the purpose of analysing its impact, technological change can be relatively easily incorporated into this framework through the notion of asset specificity. This notion refers to investments the value of which depends on the identity of one's business partners. As a consequence of such investments, one can become hostage to threats by that business partner to break up a business relationship. Specially designed governance forms need to anticipate and accommodate this situation. While creating these special governance forms is costly -- increasing unit transaction costs -- creating them can still be worthwhile since these governance forms allow for degrees of asset specificity which, in turn, may reduce transformation costs substantially. There are several types of asset specificity such as specificity of location, human skills and machinery. For example, locating one's production facilities close to one's customer's own production facilities may allow for superior logistical organization (such as Just-in-time delivery arrangements) which, however, creates a substantial degree of asset specificity since the value of that investment would be much reduced if the customer would sever the business relationship. Technological change can then be incorporated as investments in new machinery. Since initially such innovative machinery still follows proprietary or unproven designs, it may well be tailored to the specific needs of one business partner and therefore increase the de-

gree of asset specificity in a business relationship. These concepts are summarized in Figure 1 which extends and slightly modifies a figure presented in Williamson (1993).

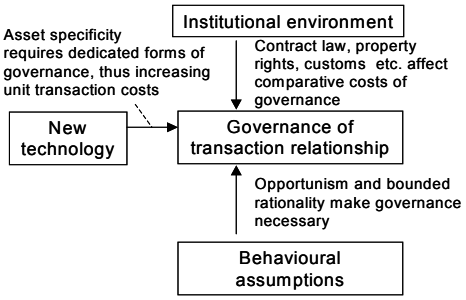


Figure 1: Technological change in a Williamsonian framework (modified from Williamson (1993) excluding feedback and indirect effects and adding the effect of new technology).

The Northian framework combines elements of neoclassical economic theory and transaction cost economics. North is not directly concerned with explaining or describing forms of governance but rather analyses development of macro entities such as whole economies/societies or sectors within one economy. In order to compare his approach to that of Williamson and to make it relevant for our analysis, it therefore needs to be 'mapped' onto the transactional level. In North's (1990) framework forms of governance are treated as behaviour of organizations on economic markets which is influenced by formal and informal institutions through affecting levels of transaction costs (organizations are the players, institutions are the rules of the game). Some organizations engage in the production of new knowledge which, if incorporated into products, becomes technological change on economic markets. In addition, some organizations also engage in activity on political markets in order to affect institutions (and thus the rules of the game played on economic markets). Whether such political activity results in transaction cost reducing or increasing changes of the institutional framework largely depends on ideology (which thus becomes a major force in explaining rise and decline of whole societies). In analogy to the entrepreneur who drives technical progress on economic markets, North (1981) also introduces the political entrepreneur who drives change on political markets (which activity is in addition to the political activity of organizations). (North has incorporated some more elements in his framework such as markets for knowledge creation which we suppress here in order to keep complexity on a level allowing for comparing the two frameworks; for a more detailed description and analysis, cf. Reimers (1995), pp. 11-18).

When mapped onto Williamson's three layer model, the role of institutions is mostly identical to that assigned by Williamson to them. They act as "shift parameters" (Williamson, 1993) affecting the costs of doing business on the economic markets, i.e. transaction costs. However, changes on this level are incorporated into the analysis and are thus given a more active role while in Williamson's framework institutional changes are treated as exogenous events. Thus, actions of organizations or entrepreneurs to affect institutions and thus indirectly behaviour on economic markets (including design of governance forms) is included in the framework. While such behaviour is affected by many factors, North assigns a prominent role to ideology which can be viewed as located on the micro-level in Williamson's three layer model (cf. Figure 2 for a summary of those elements of North's framework which are relevant from the perspective of Williamson's three layer model).

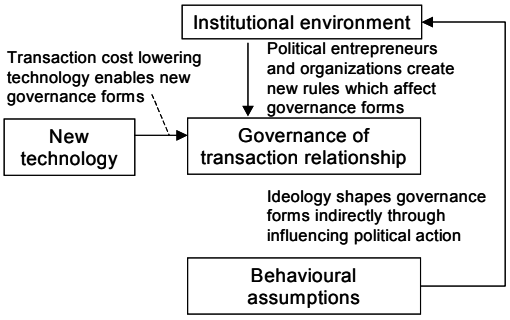


Figure 2: Technological change in a Northian framework (as mapped onto Williamson's three layer model).

The main difference between the two frameworks -- apart from the different roles new technology plays in them -- consists in the effect ascribed to behavioural assumptions. In Williamson's framework, behavioural assumptions necessitate creation of dedicated governance structures which thus take a prominent place in the framework and generally increase unit transaction costs. In North's framework, behavioural assumptions affect behaviour on political markets (which take centre stage in the framework) which then affects transaction costs on economic markets. Technological change enters through different routes too, albeit in both cases indirectly. In Williamson's framework, technology appears as a special form of asset specificity (thus, by necessity, increasing unit transaction costs); in North's framework, technological change enters as 'technical progress' created on 'knowledge markets' (given the right ideology has emerged stimulating

the demand for 'productive' knowledge; in Figure 2, this relationship between new technology and knowledge markets in not expressly modelled).

The question arises how these two explanations can be empirically tested. While it would be tempting to approach the problem in a direct manner by estimating the effect of new technology on transaction costs (whether it lowers or increases unit transaction costs), such an approach would not be acceptable from a Williamsonian perspective since by assumption -- as pointed out and criticized by North -- transaction costs cannot be measured directly according to that perspective. However, it should be possible to observe for which purpose new technology is being used, taking use as an indication for its effect (on transaction and/or transformation costs). According to a Northian interpretation, new technology which coincides with the emergence of new governance structures should be used for purposes of preparing, negotiating, or monitoring contracts while, according to a Williamsonian perspective, new technology which coincides with the creation of new governance structures should be used for transformative purposes (making production and/or distribution/logistics more efficient).

In addition, certain characteristic differences should be observed. While institutional changes (i.e. changes in the institutional environment as defined by the three layer model) are not mandatory for new technology to have its predicted effects, they can coexist with technological change. However, from a Williamsonian perspective, technological change and institutional change, should they both be observed simultaneously, would have independent effects on governance structures except for the case of institutionally mandated technological change. While institutional change may increase or decrease transaction cost levels technological change will always increase unit transaction costs. Thus, technological change will only occur if the resulting increase in unit transaction costs is accompanied by an even larger reduction in transformation costs or if changes in the institutional environment lead to the adoption of governance structures which are required to accommodate the technological change. In contrast, from a Northian perspective, one would expect political entrepreneurs or economic organizations to lobby for institutional change which would exploit the transaction cost lowering properties of new technology if such technology should become available. Thus, institutional change would have a different character in both cases. From a Williamsonian perspective, it would have the character of enforcing new governance structures (or be completely independent of technological change) while, from a Northian perspective, it would have the character of being enabled by technological change.

Finally, behavioural assumptions would play different roles in the two theoretical frameworks and thus in any explanation of empirical phenomena. Specifically, from a Williamsonian perspective, they would underlie the design of governance structures; from a Northian perspective, they would shape political behaviour by political entrepreneurs and/or economic organizations.

The discussion can be summarized by the following two propositions:

Proposition 1 (Williamsonian interpretation): Technological change which coincides with institutional change and affects the governance of transactional relationships leads to systems which increase efficiency of logistical and/or production processes. In addition, institutional change takes on the character of enforcing the new governance structure which, in turn, will reflect behavioural traits of agents.

Proposition 2 (Northian interpretation): Technological change which coincides with institutional change and affects the governance of transaction relationships leads to systems which are used for lowering relevant transaction costs. In addition, institutional change takes on the character of being enabled by the technological change while behavioural traits of agents are reflected in the institutional change.

3 DESCRIPTION OF THE CASE

3.1 Description of Data Sources and Method

We used a single case study design for our research. Use of the case study research method is justified because our questions concern the 'how' and 'why' of an organizational phenomenon which cannot be studied outside its real-life context and which involves interaction of a large number of variables (Yin, 2003). Use of a single case design is justified because of the uniqueness of the case -- massive institutional change on an industry level coinciding with technological change -- and the difficulties of obtaining empirical data in that context (ibid.). Specifically, the healthcare industry in China is exposed to an intense public debate about its practices because of a widespread dissatisfaction with healthcare services and the difficult situation of central

government in this discussion. For these reasons, it is very difficult and time-consuming to gain access to key informants in the industry. We have therefore focused our research on the case of drug distribution in Beijing.

In total, 12 interviews were conducted with 15 informants between September 2004 and May 2007. Informants represent manufacturers, wholesalers, e-commerce intermediaries, a so-called bidding centre (a government agency), the Ministry of Health, and the key person driving the introduction of e-commerce into drug distribution in China. That person was interviewed four times over the research period, allowing us to follow the evolution of the technological and institutional change over three years. In addition, documented material -- mostly in the form of Chinese websites -- was used for supplementing our data.

All information presented in the case description below has been triangulated by at least two interview sources except in cases where informants represented the subjective view of their organizations; such instances are explicitly in1dicated in the following description when they occur.

3.2 The Problem which Triggered Institutional Action

The reform of economic structures in China started in the late 1970s has not only led to the emergence of business organizations and economic markets but also deteriorated existing economic organizational structures. One characteristic of these previously dominating structures was the tight integration of work organizations and social services such as housing, education and healthcare. These integrated units -- called *dan wei* -- were not only internally integrated but, to a large extent, externally insulated. Workers would seldom leave the compounds on which all facilities required for everyday life existed. The main connection with the economic environment consisted of flows of intermediate goods among these organizations (Walder, 2000). Thus, the functional separation that came with the emergence of dedicated business organizations -- as opposed to these integrated work organizations -- implied that social services would have to either be provided as a commercial service as well or by government. To some extent, both of these directions were pursued, especially with regard to healthcare. Specifically, while, through a reform of the health insurance system in 1998, all workers in cities are covered by a governmental insurance system, rural families receive practically no any health insurance coverage (cf. Dou, 2003, and IMS Health, 2004). Most medical expenses need to be paid out of pocket by rural families (ibid.). At the same time, governmental health insurance for urban workers covers only basic services so that a large number of privately-based insurance schemes has sprung to live covering additional health risks (ibid.).

While this situation was not satisfactory for most people and organizations involved in healthcare, it continued to function to the extent that healthcare costs could be kept low. The healthcare system started to be defunct, however, once the drug prices started to increase significantly. While one cause of the rise in drug prices was the entering of multinational pharmaceutical firms into the Chinese market and the accompanying rise of branded drugs -- alongside the much cheaper so-called ethnic drugs, i.e. drugs based on traditional Chinese medicine -- the root cause for this development was the chronic under-financing of hospitals. In order to survive in the new economic environment, hospitals took to earning most of their income (on average 80%) through the selling of drugs which naturally created incentives to sell expensive drugs with high margins.

Central government initiated several institutional measures to mitigate the situation. For example, it kept prodding provincial and local governments to improve healthcare provisioning and to develop insurance schemes for the rural population. It also instituted that all business organizations operating in the distribution of drugs had to be certified by the year 2004. The reason for this measure was an intention to cut down on the huge number of distributors, wholesalers and other intermediaries which, around the year 2002, was estimated to be between 16 and 17 thousand (Dou, 2003). This large number of intermediaries in the distribution of drugs was supposed to create inefficiencies through fragmentation (lack of economies of scale) and multiple mark-ups (each intermediary would add a mark-up to the price). Moreover, central government required hospitals to separate their internal pharmacy accounts from their other accounting processes in order to increase transparency regarding the extent to which hospitals financed themselves through the sale of drugs; a second institutional reform concerned the introduction of a centralized bidding process through which hospitals were expected to purchase drugs. There has also been some efforts to promote the development of an independent retail pharmacy sector because it was assumed that through this process the monopoly power hospitals traditionally held over the sale of drugs could be broken or at least diminished.

All these measures are very recent, beginning in the year 2000, and government is continuing to experiment with new approaches. However, government is severely restricted in enforcing its policies

for two reasons. First, it depends upon the services of hospitals, a fact which came to light during the SARS epidemic in 2003. Therefore, government cannot afford to let a large number of hospitals go out of business. Second, government is not a unified force but internally highly fragmented along vertical and horizontal lines. Specifically, regulatory and administrative powers regarding the healthcare sector were, in 2002, distributed across nine governmental agencies and ministries (Dou, 2003) some of which were later merged. Vertically, government power is spread across central and provincial governments, for example with respect to inspecting and certifying drug manufacturing.

In the following, we will focus on the introduction of the centralized purchasing process since this was the main force shaping the development of electronic commerce systems in drug distribution but also consider the effects of the other measures since they are all interdependent.

3.3 An Early Experiment

In order to curb corruption related to the problems outlined above, the provincial government in Henan province centralized all drug procurement related to hospital demand in 1993. Resistance by hospitals to this measure, however, lead to the discontinuation of the practice two years later. The person in charge of implementing this measure, Mr. Li, then visited the US in order to study drug distribution there and became convinced that centralization of drug procurement was the "direction for the future of China" but that this was only possible by using e-commerce. After two years of preparation, an e-commerce system for procuring drugs commenced operations, initially with good results -- according to Mr. Li -- but which was closed down after just half a year of operations upon being declared illegal by central government under then premier minister Mr. Zhu Rongji because business operations had to be separated from their regulation. The system was then sold to a private company -- Haihong -- which would re-launch it in Henan and introduce it to several other provinces, among them Guangdong, Hainan, and Beijing.

3.4 How the System is used in Beijing

In Beijing, a centralized, e-commerce-based procurement system for drugs began operations in 2004. However, institutional reforms preparing for that system go back to the year 2000 when the 'Beijing Bidding Centre' was set up as a joint effort by nine government agencies involved in the regulation of drug distribution to hospitals. These agencies include, for example, the Beijing Health Bureau (which has a role similar to a national Ministry of Health albeit on a provincial level), the Beijing Price Bureau and the Beijing Traditional Chinese Medicine Bureau. Based on experiences with similar systems in other provinces, most of them supported by Haihong's software and services, a process for drug procurement was established; while the core elements of this process are similar across the whole country some elements show distinct characteristics distinguishing the process from those implemented in other provinces. In general, drug distribution is still a highly localized business; distributors who are fierce competitors in one province may therefore be business partners in 1another province. The following account focuses on the practices associated with the e-commerce system in Beijing.

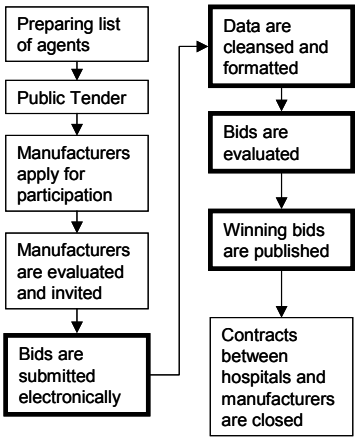


Figure 3: The bidding process (bold boxes indicate steps supported by the e-commerce system).

The bidding centre usually initiates a bidding process once per year. The process steps are depicted in Figure 3. The core step consists of evaluating competing bids by manufacturers on individual pharmaceutical agents (chemical substances) according to multiple criteria, including price and service quality of distributors (which must be assigned by manufacturers in advance and which often take over the paperwork associated with participation in the bidding process). The list of agents comprises about 15 thousand items on which manufacturers can bid. Evaluation is done by a group of experts for each of the province's -- Beijing is a province as well as a city -- six hospital groups which have been set up according to certain differences in their demand for drugs. Once bids have been selected, hospitals are required to place purchase orders for drugs only among the winning bids. The main purpose of this process is to ensure that hospitals use high quality drugs while controlling drug prices.

The bidding process is facilitated by a number of intermediaries which have been certified for that purpose. While seven intermediaries have received such certificates, only three are active. Each hospital group selects one of them to help them with operating the bidding process. The largest among them -- serving four hospital groups -- is Haihong. The requirements according to the certification process also include ownership and operation of an e-commerce system. Regarding the bidding process itself, that system collects all documents which need to accompany a bid and which have to be submitted electronically. The intermediaries then pass these data -- after some data cleansing and format adjustments -- on to the bidding centre. Once winning bids have been selected, the results are published on the e-commerce systems.

The more important role which these e-commerce systems play, however, concerns the ordering process. Hospitals are required -- according to stipulations by the Beijing Health Bureau -- to submit orders through these systems. For that purpose, they log onto the system over a web interface and enter their orders directly into the system. Distributors then download order data from the system -- also by logging onto the system through a web interface -- and hospitals are automatically informed that distributors have downloaded order data. However, distributors cannot confirm or change purchase orders.

The express purpose of this use of e-commerce systems is to facilitate the monitoring of compliance with the rules of the bidding process. Specifically, the bidding centre is charged with the task of monitoring hospital purchasing activity in order to ensure that hospitals only buy 'from the list', i.e. do not circumvent the drugs which have been selected in the evaluation process.

The percentage of drugs sold/procured through these systems has increased continuously since their inception in 2004 and was estimated to be close to 100% in 2006. Use of the system is accredited with having caused a significant drop in the number of distributors in Beijing which fell from around 200 in 2004 to about 120 at the beginning of 2007.

The e-commerce system replaces a practice in which distributors took drug orders from hospitals by phone (orders by distributors to manufacturers continue to be placed over the telephone or by fax). The services of intermediaries complement the e-commerce system's functionality, mostly by offering a 'screening service'; if a distributor does not respond to an order, the intermediary will help the hospital to procure the drugs through other channels. The intermediaries also improve the efficiency of the ordering process by harmonizing data. For example, hospitals often use internal codes to identify drugs. These codes are matched to standard drug identifiers defined by the China Food and Drug Administration (FDA) so that distributors can use the FDA codes for their internal processes rather than having to cope with multiple proprietary codes used by hospitals.

While some of the data used for the ordering and for the bidding processes are identical, the systems are separated (including separate databases) because they are regulated by different government agencies (as a side-effect of this separation, the ordering system could also be used by other organizations who do not have to participate in the bidding process such as independent pharmacies; because of capacity problems, however, use of the system is currently limited to hospitals). Operational efficiencies of using the e-commerce system for supporting the bidding process are minor as compared to the ordering process. The main benefits concern ease of selecting and evaluating drugs which facilitates the work of the expert group who selects bids. Again, this requires harmonization of data supplied by manufacturers (or distributors acting on behalf of manufacturers) which is done by the intermediaries.

The vision of Mr. Li, who continues to advise Haihong regarding further development of its e-commerce system, is to provide a comprehensive platform for managing the whole drug supply chain from manufacturers to hospitals. However, currently only the order process is supported by the system and even this support is rather limited (as evidenced by a lack an order confirmation or change function). For example, it was intended to use the system for enabling zero-inventories in hospital pharmacies. However, hospitals were not interested in such a capability because they do not have to pay manufacturers for unsold inventories. While Haihong's system was the first to be developed, the systems of all three intermediaries are rather similar in terms of functionality and capacity and do not constitute a main competitive differentiator for them.

Thus, while the operational scope of the three e-commerce systems is rather narrow, they are indeed used for improving operational efficiency of the ordering process. For example, while, in 2004 the association of drug wholesalers and pharmacies received broad support by its constituency when submitting a petition to government objecting introduction of the e-commerce-based distribution system, that support is waning as large distributors (wholesalers) are discovering operational benefits of the system and because the system has led to a consolidation of the industry favouring the larger players.

Hospitals have to contract with the intermediary as well as with manufacturers. Each hospital group

selects one intermediary which, however, charges fees to manufacturers as stipulated by the Beijing Health Bureau. The licenses of intermediaries need to be renewed each year but the relationship between a hospital group and an intermediary tends to be stable and long-term. Upon conclusion of the bidding process, hospitals have to contract with manufacturers that have succeeded in the bidding process. Occasionally, hospitals also negotiate with manufacturers again before placing orders on the e-commerce system in order to receive discounts. In the evaluation process, it is also possible that the bidding centre negotiates with manufacturers who have participated in the bidding process. This occurs when for a specific agent (chemical substance) only one bid has been submitted. In addition, manufacturers use sales agents and other wholesalers to market and also sell their drugs to hospitals. However, these agents, who may receive the drugs at a discounted price, still have to use the licensed distributors for delivering the drugs to hospitals.

The specific governance structure used for operating e-commerce systems through intermediaries is justified by two rationales. First, the main benefit of the system is seen in the ease of monitoring compliance with the bidding rules. This is considered crucial since hospitals have very strong incentives to circumvent these rules in order to increase their income through the sale of branded and therefore high-margin drugs. Second, Mr. Li argues that e-commerce in China is only viable if facilitated by third parties. He points to an effort in Shanghai where wholesalers have tried to build e-commerce systems to directly connect with hospitals. These efforts have failed because of the fragmented market structure (hospitals typically deal with around 30 different distributors), fierce competition among distributors (forestalling cooperation among them) and low trust among all parties.

The e-commerce systems, however, are not used for the purpose of monitoring compliance with the bidding rules. Specifically, the bidding centre does not make any use of its ability to log onto the systems in order to check hospitals' compliance with the bidding rules (as reflected in their ordering behaviour). While the bidding centre claims that 50-60% of all purchasing transactions are reported to it by intermediaries, it turns out that this feedback is based on aggregated data which are provided by intermediaries to the bidding centre on paper and this only upon request which occurs infrequently and usually only once per year. The intermediaries suggest that the bidding centre lacks the technical skills required for making sense of the data provided by the systems directly. The bidding centre itself indicates that its ability to sanction hospitals (through exposing non-compliance) is rather limited because it is difficult to tell violations of rules from "market behaviour", a view shared by Mr. Li. It was also frequently mentioned that hospitals often have sufficient "market power" to resist any sanctioning efforts. Mr. Li cited yet another reason for the failure of the bidding centre to directly use the e-commerce systems to monitor purchasing behaviour of hospitals. The data in the systems cannot be easily analysed because of a lack of data standards (apart from the use of proprietary product codes by hospitals mentioned above, other data such as names of manufacturers are not standardized either). Moreover, some hospitals ask intermediaries to provide them with so-called 'soft systems' for their data input which are tweaked so as to make it even more difficult to monitor their purchasing behaviour.

4 CASE ANALYSIS

Referring to the two propositions developed in the theory section, it becomes clear that the e-commerce systems described above have been developed with a Northian intention. The main idea of Mr. Li -- who can be viewed as the political entrepreneur described by North -- was that centralization of the procurement process -- i.e. a specific instance of institutional change -- was only possible if accompanied by an e-commerce system. He assumed that electronic commerce would facilitate the monitoring of hospital drug purchasing behaviour which was a central element in the introduction of the bidding process. These costs are an instance of transaction costs since they concern the monitoring of a contract, albeit one imposed on the participants in the market. Thus, the e-commerce system's ability to reduce transaction costs would be exploited for the purpose of institutional reform. Also, the organizations active on the economic market -- mostly distributors and hospitals -- tried to prevent the institutional reform through lobbying activity.

The bidding process centralized all negotiations between hospitals and manufacturers while the e-commerce system centralized ordering and delivery (for a small group of hospitals in Beijing, government has actually taken over all procurement activities, thus reversing the earlier decree by central government to separate regulation and business activity). Also, the very idea of solving the problems characterizing drug distribution through taking over

central control of the process seems to be in line with the overall ideology of economic policy in China. Thus, all elements suggested by a Northian interpretation of the interaction between institutional and technological change -- technology-enabled institutional change, activity on political markets to bring about or prevent the institutional change, a 'national' characteristic trait of the intended institutional reform -- seem to be in place.

However, the e-commerce system is not used for the intended purpose (facilitating institutional change through reducing the costs of monitoring compliance with the new rules); yet, the main actors in the industry continue to offer the rationale of facilitating monitoring of hospital drug purchasing behaviour as an explanation for continued use of the e-commerce systems. While there are several reasons for the factual avoidance of using the system for monitoring hospital transactions, including bargaining power of hospitals and lack of alternative funding schemes, the fact that the e-commerce systems are not used for their intended purpose allows us to reject the Northian interpretation of this case. The new institutional process -- the bidding process -- was not enabled by a reduction of transaction costs -- in this case monitoring costs -- through use of the e-commerce system; continued existence of the new institutional arrangement can only be explained by government's use of sanctions and administrative force. In addition, significant violation of the rules stipulated by the bidding process was tolerated by government agencies which may have reduced resistance to the new institutional order.

Regarding a Williamsonian interpretation, it seems reasonable to argue that the e-commerce systems did not reduce transaction costs but probably contributed to increased efficiency of the ordering process. For example, distributors do not have to handle multiple systems for identifying drugs through proprietary product codes while hospitals can place orders to distributors in one process rather than sending individual orders to each distributor separately. (Note that costs associated with placing orders are not transaction costs since they neither concern the costs of negotiating contracts -- this has been done through the bidding process resulting in bilateral contracts between hospitals and manufacturers -- nor do they constitute costs of monitoring contracts; if the supply chain would be controlled by one organization, it would still be necessary to create (internal) orders for replenishing inventories in hospital pharmacies.)

On the other hand, transaction costs associated with the new governance structure may well have been increased because the bidding process introduced new costs into the distribution system; e.g., the expert groups have to evaluate bids on 15 thousand agents (chemical substances) by multiple suppliers. While the e-commerce systems provide some support for this process, the main work is manual. In addition, multiple other channels for negotiating between hospitals and manufacturers continue to be used. Finally, bilateral contracting between hospitals and manufacturers is still necessary, although these contracts are based on the prices quoted in the bidding process. In any case, hospitals would prefer to bilaterally negotiate with manufacturers even so they do not have to pay for the work of the bidding centre (which is completely funded by the Beijing government). Also, mandatory use of the new intermediaries did impose some new costs on the distribution system. Presently, these are carried by the manufacturers (who must -- by law -- be charged for the intermediaries' services -- although they serve the hospital groups). It was anticipated that hospitals would effectively oppose any efforts to make them pay for the services of intermediaries, e.g. by asking manufacturers to take over these fees).

The governance structure used for facilitating set-up and operation of e-commerce systems also seems to display typical 'Chinese characteristics' as these systems are developed and operated by third parties rather than within a bilateral governance structure typical for Western EDI-based models of e-commerce. Thus, all elements suggested by a Williamsonian interpretation can be identified in the way these systems are actually used. The governance structure underlying operation of the e-commerce systems is likely to have increased overall transaction costs in the distribution of drugs between distributors and hospitals. However, operational efficiencies in the ordering process could (partly) compensate for these increased transaction costs. As the operational efficiencies become clearer, resistance towards using the e-commerce systems wanes (this fading resistance seems also to be related to recognition among the main players that the systems are not used for monitoring their market behaviour).

However, because of a lack of hard data, a Williamsonian interpretation cannot be proven. Moreover, this lack of hard data could well be implied in the approach itself, a problem which motivated the critique of North and Wallis in the first place. Yet, we find that the Northian interpretation explains the motives behind the political initiative to create a centralized, e-commerce-based drug procurement process while actual use of the system contradicts such an interpretation.

5 CONCLUSIONS

In exploring two alternative transaction cost-based explanations of the interaction of institutional and technological change in the case of China's drug distribution system, we have found evidence that a Northian interpretation -- which suggests that transaction cost reducing technological change enables institutional change -- does not apply to the actual practice of distributing drugs based on new institutional rules and an e-commerce system but does accurately describe intentions and political behaviour which has resulted in these new arrangements. Actual use of the e-commerce system and its governance structure seems to be more in line with a Williamsonian interpretation which suggests that institutions and governance structures need to be adapted to enable implementation of new technology in organizational systems, resulting in increased operational efficiency, albeit at the expense of increased transaction costs.

We therefore conclude that development and use of new technology as well as intentions associated with an institutional change and the practices based on these changes may well follow different rationales. More generally, while a certain *perceived* interaction between technological and institutional change may motivate these changes, the practices which result from these changes could follow a different logic of interaction. This result is reminiscent of early formulations of neo-institutional theory (Meyer and Rowan, 1977). These formulations suggested that organizations maintain a 'gap' between their internal operational processes and their external, symbolic justification of these processes. While these early propositions have been strongly criticized by institutional theorists themselves (Powell, 1991), our results suggest that such a gap could be the result of an implementation process in which the rationales that motivated technological and institutional change continue to exist as a 'distant echo' in actual practices. However, actual practices have adapted to a different rationale.

While our results are not sufficient to accept or reject one transaction cost-theoretic interpretation over the other, they suggest that further exploration of these two explanations in settings which study the interaction of technological and institutional change on an intermediate level between that of individual organizations and whole economies is a worthwhile effort. Moreover, our study shows that the two explanations could account for different phases in the evolution of that interaction, a result which was not expected at the outset of this study.

ACKNOWLEDGEMENTS

The first author would like to acknowledge support by the Deutsche Forschungsgemeinschaft (grant number 1328/2-2); the second author would like to acknowledge support by the China Natural National Science Foundation (grant numbers 70621061 and 70672007) and the China MOE Project of the Key Research Institute of Humanity and Social Sciences at Universities (number #06JJD630014).

REFERENCES

Dou, W. (2003). *The Chinese Pharmaceutical and Biotechnology Market Guide 2003*. SCRIP Report No. BS1210, 28 March 2003, Richmond/Surrey: PJB Publications.

IMS Health (2004). *IMS Market Prognosis Australasia 2004-2008 – China*. IMS Health, London, March 2004.

Meyer, J. W., & Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology*, *83*, 440-463.

North, D. C. (1981). *Structure and Change in Economic History*. N.Y., London: W. W. Norton.

North, D. C. (1990). *Institutions, Institutional Change, and Economic Performance*. Cambridge: Cambridge University Press.

North, D. C., & Wallis, J. J. (1994). Integrating Institutional Change and Technical Change in Economic History -- A Transaction Cost Approach. *Journal of Institutional and Theoretical Economics*, *150*, 609-624.

Powell, W. W. (1991). Expanding the Scope of Institutional Analysis. In: W.W. Powell & P.J. DiMaggio (eds.), *The New Institutionalism in Organizational Analysis* (pp. 183-203). Chicago et al.: University of Chicago Press.

Reimers, K. (1995). *Normungsprozesse – Eine Transaktionskostentheoretische Analyse*. Wiesbaden: Gabler Verlag (Neue betriebswirtschaftliche Forschung Bd. 146).

Von Hippel, E. (1978). A Customer-active Paradigm for Industrial Product Idea Generation. *Research Policy*, *7*, 240-266.

Walder, A. G. (2000). China's Transitional Economy. In: J.T. Li et al. (eds.), *Management and Organizations in the Chinese Context* (pp. 63-83). Houndsmills et al.: Macmillan,.

Wallis, J. J., & Douglass C. N. (1986). Measuring the Transaction Sector in the American Economy 1870 - 1970. In: S.L. Engerman & R.E. Gallman (eds.), *Long Term Factors in American Economic Growth*, Chicago: University of Chicago Press.

Williamson, O. E. (1993). Transaction Cost Economics and Organization Theory. *Industrial and Corporate Change*, *2*, 107-156.

DIABETES SCREENING DATABASE
A Comprehensive Electronic Patient Record for Global Risk Assessment in a Rural Community

E. Pecoul [1] and H. F. Jelinek [2]
[1] *IUP Génie Physiologique et Informatique, Faculté des Sciences Fondamentales Appliquées*
Université de POITIERS, France
[2] *School of Community Health, Charles Sturt University, Albury, Australia*
epecoul@gphy.campus.univ-poitiers.fr, hjelinek@csu.edu.au

Keywords: Diabetes Screening, Access database, interdisciplinary health care.

Abstract: Interprofessional health care is becoming more prevalent with an increase in chronic diseases such as diabetes and cardiovascular disease. In addition preventative models often require large numbers of risk factors for identification of preclinical cases. CSU has established a diabetes screening clinic augmented by an ACCESS database. The novelty of our work is that the ACCESS database integrates into the public health sector and provides a more comprehensive review of health/disease indicators. Information on traditional health indicators in addition to autonomic nervous system function tests, fundus examination and foot assessment results as well as pro-inflammatory, pro-coagulation and antioxidant biochemistry can be added by the university-based screening clinic as well as by diverse primary health care practitioners and specialists that would otherwise not have access to this detailed information for patient assessment and treatment. Our results are in favour of this interdisciplinary database indicating that over one year we have identified 16.2% of people with no previous medical condition to have pre-diabetes, 2% had retinal disease and 21.5% had foot problems. Moderate to severe ECG anomalies were identified in 19.3% of the participants. Of these, 68.8% were either commenced on treatment, had their treatment changed or received surgery. Our results indicate that a comprehensive EPR manager as part of an interdisciplinary health screening initiative is able to track people that require intervention but were missed in the current public health system as implemented in our community.

1 INTRODUCTION

For optimal health care, providers need the necessary information when they give treatment. Health care provider and patient information and decision support needs can be satisfied if primary care providers use electronic patient records that are comprehensive and provide the necessary information to the general practitioner or specialist at the time of consultation (EPRs) (Bates et al., 2003). Often electronic records may indicate that a patient has attended a specialist for an assessment but the results may not be available to the general practitioner nor to other primary health care providers and specialists. In addition many rural areas experience a lack in specialists and allied health professionals as well as general practitioners. Electronic patient records (EPRs) provide a powerful opportunity for health assessments yob re integrated as part of an integrated interdisciplinary primary health network. Universities with additional resources in rural communities can also to support this initiative by establishing screening clinics as part of their health professional courses. University-based teaching and research can provide screening/assessment results especially for tests not routinely carried out in the community due to lack of available services. With additional risk assessment results during consultation, the GP's EPR is valuable in identifying patients at risk for undiagnosed type 2 diabetes, cardiovascular disease or other illness. Previous research on use of electronic patient records (EPR) indicated that general practitioners are able to utilise information on additional risk factor assessment from the EPR during regular consultation (Klein Woolthuis et al., 2007). EPR recall systems based on HbA1c levels showed that the median practice-specific proportions of patients

with HbA1c recorded annually increased from a median 65% to 77% when using an EPR (Mohiddin et al., 2006). An Australian study conducted by Burns et al. using the CARDIAB®^{TM} register recall system evaluated the process of care of patients for nine parameters critical to quality of care in diabetes: glycaemic control (HbA1c), blood pressure, body mass index, total cholesterol levels, triglyceride levels, high density lipoprotein levels (HDL), microalbumin level, foot checks and eye checks. They demonstrated the potential of CARDIAB®^{TM} to provide support to general practice in the management of diabetes (Burns et al., 2004).

Divisions of General Practice, universities and other primary health care providers have a key role to play in supporting general practitioners (GPs) to provide proactive, preventive care for their patients with cardiovascular disease (CVD) and diabetes (Penn et al., 2004). They can achieve this by providing them with global risk marker information and health assessment results. Risk stratification is currently based on the Framingham cardiovascular risk equation and therefore include factors such as blood pressure, diabetes, age, gender and HDL (Sheridan et al., 2003). However research has suggested that a more global approach to identification of diabetes and cardiovascular disease may be appropriate (Michos et al., 2006). In addition the risk status of persons without CHD varies greatly, and this variability mandates a range in the intensity of interventions. Effective primary prevention thus requires an assessment of risk to categorize patients for selection of appropriate interventions and a more comprehensive screening /health assessment that includes additional factors provided by combining the information obtained from an interdisciplinary health care system and available via an EPR (Jelinek et al., 2006, Woodward et al., 2006).

The major and independent risk factors for CHD are cigarette smoking, elevated blood pressure, elevated serum total cholesterol and low-density lipoprotein cholesterol (LDL-C), low serum high-density lipoprotein cholesterol (HDL-C), diabetes mellitus, and advancing age. Other factors are associated with increased risk for CHD. These include obesity, physical inactivity, elevated homocysteine, pro-coagulation and pro-inflammatory factors. An assessment of global risk based on the summation of all major risk factors can be clinically useful for in identification of high-risk patients who require immediate attention and intervention and changes in intervention based on the global risk estimate (Grundy et al., 1999).

Health assessment and determination of risk factors included in a global risk equation is often not possible in rural communities where there is a lack of general practitioners, allied health practitioners and specialists. University-based teaching and research has the opportunity to integrate with the public health system (Jelinek et al., 2006). An electronic patient record system that spans from university-based teaching and research to public health care and that incorporates data obtained from an interprofessional health care system is required.

The EPR must integrate medical, nursing, and allied health worker data as well as the results from pathology laboratories that may be shared by the health care team. The EPM needs to be secure, provide relevant template forms, incorporate a recall system and specific reports, and a means of tracking interventions by diverse health care professionals. The majority of databases meet this criterion but are mainly specific for certain medical specialities such as general practice. The DiScRi database discussed here extends these fundamental principles to address the needs of an interdisciplinary health care team with emphasis on identification and follow-up of people with and without diabetes or cardiovascular disease as well as identification of those at risk of diabetes and cardiovascular disease based on a global risk assessment. Information of outcomes following referral to general practitioners, allied health and medical specialists provides information to of global health care and the status of the patient. The database allows access to all biochemistry and pathology results as well as the retinal photography, 12-lead ECG and Doppler peripheral vessel blood flow velocity traces that is collected by diverse health care professionals that would otherwise not be available in a coherent form.

2 METHODS

2.1 Database

The Diabetes Screening Research Initiative (DiScRi) application was designed using Microsoft ACCESS (under Microsoft Windows), SQL (Select Query Language) and VBA (Microsoft Visual Basic) languages. The CDM (Conceptual Data Model) and LDM (Logical Data Model) were designed using MERISE theories. The database contained 22 entities and 153 attributes (Table 1).

Accurate information is available for demographics, medical history, biochemistry, urine analysis as well as lifestyle, medication, foot, eye

and heart health including antioxidant levels, pro-inflammatory and pro-coagulation factors for a patient. There is also a direct access to files such as retinal and ECG pictures, and a resume of most important health information (BP, age, cholesterol, diabetes status, medication).

The body-mass index (BMI), the ankle-brachial pressure index (ABPI) for assessment of peripheral vascular disease, the Ewing score, which indicates presence and severity of cardiac autonomic neuropathy as well as absolute 5 year CVD risk are automatically calculated. Cover letters for the general practitioner or specialist, if required, are automatically generated with a link between Microsoft Access and Microsoft Word (an update in the database will update the Word document) and can be sent by e-mail (using Word mail function), or printed for mailing and/or storage as hard copy.

2.2 Screening Participants

To assess the utility of the DiScRi database, patients were recruited through the media and the research had Charles Sturt University Human Ethics approval. All participants had their height, weight and waist circumference measured. A urine sample was analysed and a blood sample used to determine levels of glucose, electrolyte, cholesterol, HbA1c, antioxidants, pro inflammatory and pro-coagulation markers. Retinal photographs of the posterior pole are used to determine presence of diabetic retinopathy and risk of CVD. Heart function was determined by 12-lead ECG including calculation of corrected QT interval (QTc), QT dispersion (QTd), frequency and time domain analysis results. The latter are indicators of preclinical cardiac autonomic neuropathy (CAN). CAN is also assessed using the Ewing battery. Foot health assessment includes ABPI, monofilament for peripheral neuropathy (PN), ankle and knee reflexes, and muscle tone. Thus the university screening simulated an interdisciplinary health provision system.

3 RESULTS

3.1 Database Characteristics

From the introductory screen, the database allows access the demographic data, clinical history, follow-up, biochemistry, eye, heart and foot assessment results.

Figure 1 shows an example of a test recording screen. For the ECG analysis, the 12-lead ECG trace can also be accessed. The 3-lead recording indicates the results for the assessment of cardiac autonomic neuropathy based on the Ewing score.

All data is divided into tables that can be

Table 1: Entities contained in the DiScRi database.

Entities	# values	Examples
GP	4	Name, address, telephone, email
Patient	5	Name, address, telephone, DOB
Analysis	2	
Medications	2	Current medications, change medications
Clinical history	5	Diabetes, HT, CVD, medical information
Foot history	3	Presence of ulcers, general pain, numbness
Heart history	10	Angina, fainting, heart attack, stroke, chest pain
Eye history	2	Glaucoma, cataract, diabetic retinopathy
Other history	6	Kidney, bladder, GIT
Lifestyle	7	Smoking, alcohol, family history for diabetes/CVD, diet, exercise
Consultation history	6	Attendance to health care professionals; details of consultations
3lead ECG Ewing	7	Lying to standing HRV change
3lead ECG HRV	17	Frequency and time domain results for HRV
12lead ECG	8	HR , QTc, QTd, interpretation,
Neuropathy	7	Monofilament, ABPI,
BMI	4	Height, weight, waist circumference
ABPI	5	Systolic blood pressure at arm and ankle
Glucose	5	time since last meal, glucose level at screening
BP	5	Lying and standing BP
Urine	12	Glucose, protein, albumin
Biochemistry	26	Cholesterol, antioxidants, pro-inflammatory, pro-coagulation markers
Eye analysis	5	Eye pathology, AVR diagnosis, photograph

separately accessed, However for a quick overview during a consultation, the global information page indicates the date of last visit and includes the possibility of accessing the eye photographs of the retinal posterior pole, the 12-lead ECG trace and information on glucose and cholesterol levels as well as blood pressure and BMI (Figure 2). From this page the user can also move to any of the analysis for specific and more comprehensive results.

An important addition to this database is the opportunity for detail follow-up provided by diverse health care practitioners as shown in Figure 3.

3.2 Screening Outcomes

A review of the outcomes for the university-based simulation study that measured health/disease indicators indicated that in this rural area the prevalence is worse compared to the national average for all tests in the diabetes cohort and elevated for the ABPI and ECG in the non diabetes group (Table 2).

Table 2: Prevalence of pathology in a rural cohort determined through a university-based screening initiative.

Measure*	% Non diabetes	% National Average
PN	8.13	19.4 (foot ulcer)
ABPI	11.6	unknown
BMI	25	20 (>30kg/m2)
HT	28	30 (140/90)
ECG	13	12(diab) / 3(control)

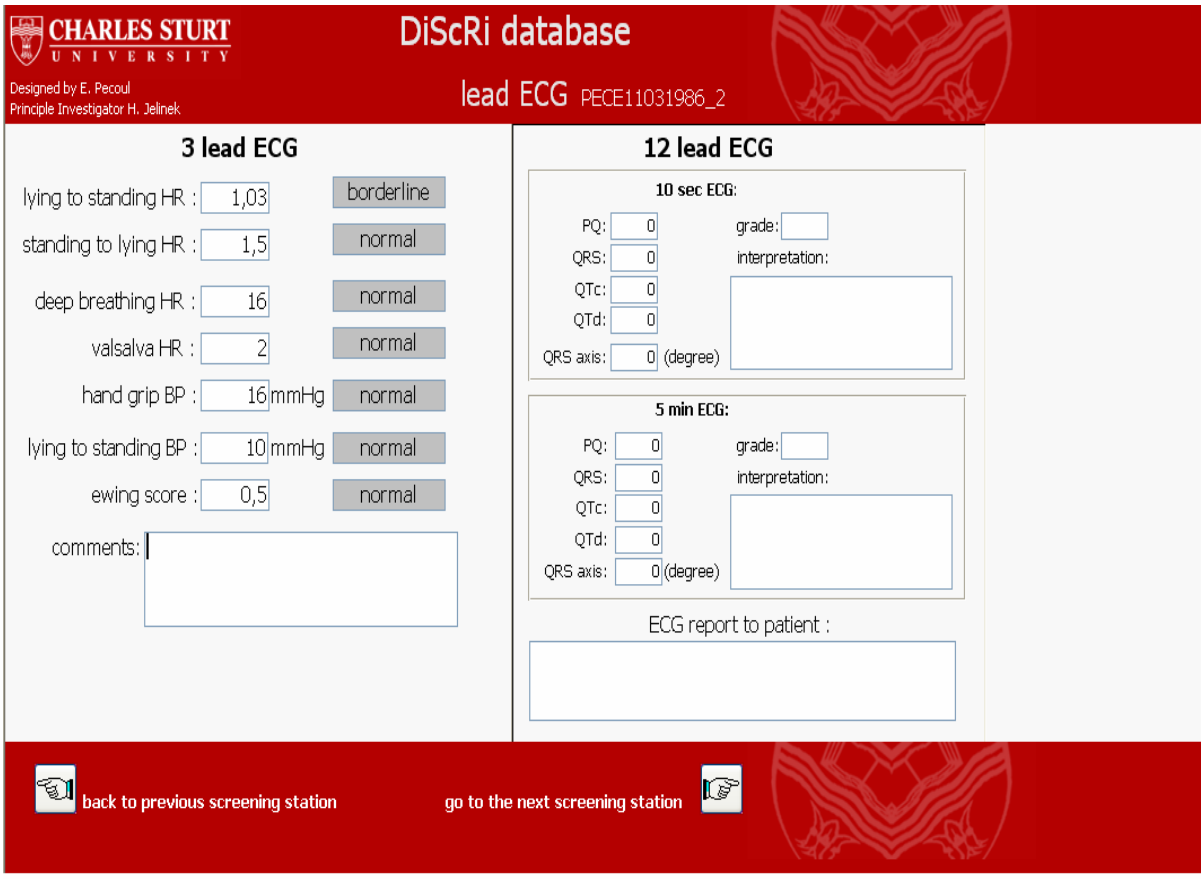


Figure 1: ECG assessment form.

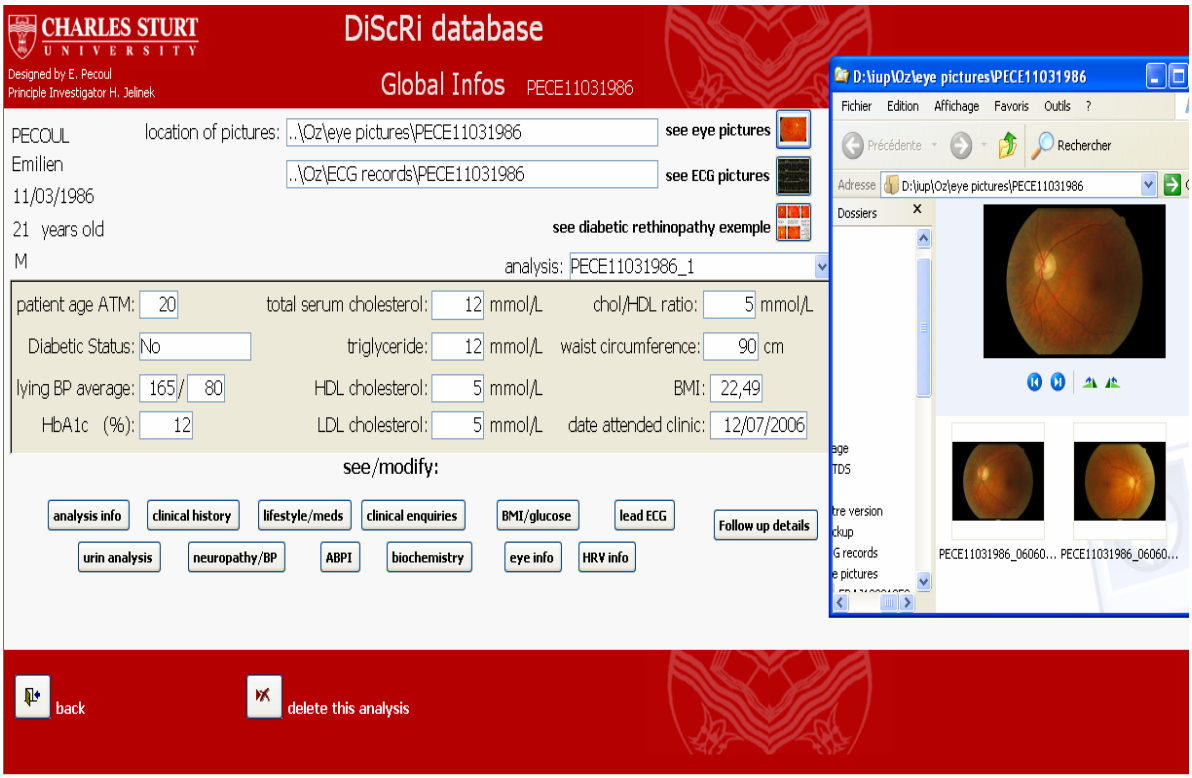


Figure 2: Global information screen with eye photograph.

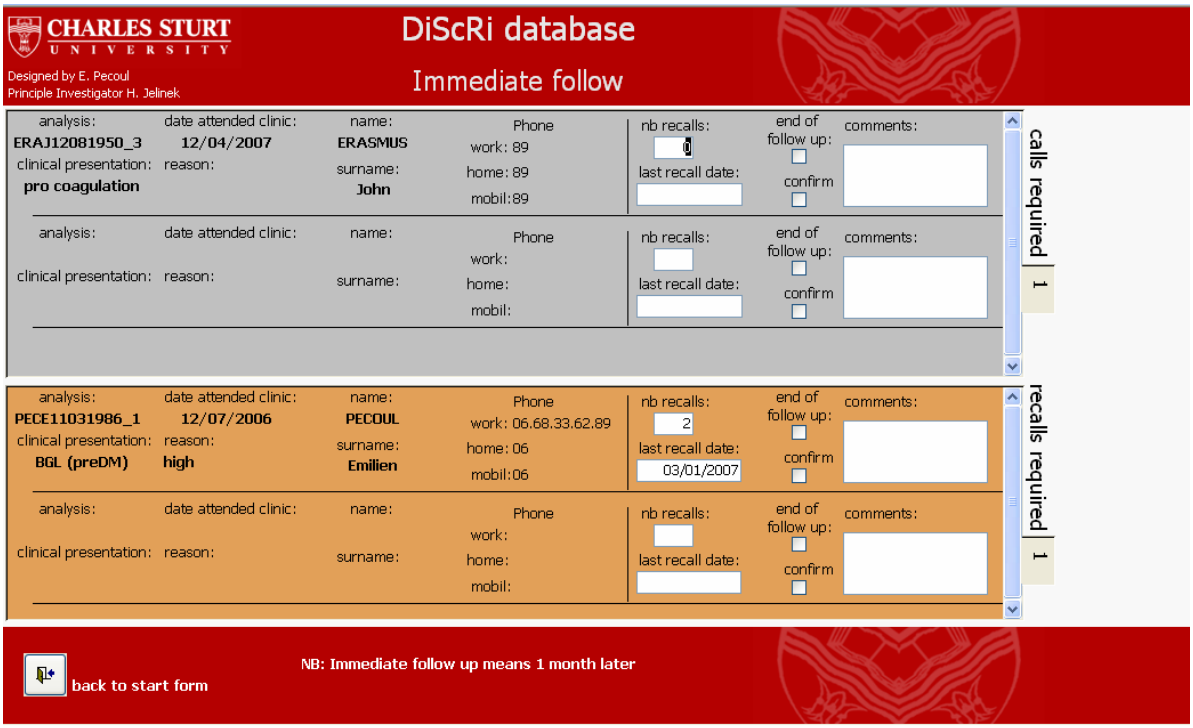


Figure 3: Recall screen.

4 DISCUSSION

In Australia, general practitioners play an important role in the management of diabetes. This involves careful monitoring of behavioural risk factors, blood pressure, glycaemic control and lipids; early detection of complications; and management according to evidence-based guidelines (NHMRC, 2004, Newnham et al., 2004). However, a comprehensive health review can only be provided by an interdisciplinary health care team. With the rapid advances in information technology in the last decade, various diabetes information systems have evolved in different parts of the world. Availability of new technologies and information systems for monitoring and treating diabetes is critical to achieving recommended metabolic control. A comprehensive EPR that includes data for global risk assessment and patient review and provides mechanisms for practitioners to gain information on performance and results from a diverse primary health care team that would otherwise not be integrated is essential for evidence-based practice and improvement in health care provision.(Joshy and Simmons, 2006) The DiScRi study indicates that an EPR that incorporates information across the health care sector arising from annual consultations in the university setting such as test results and outcome of referrals, as well as lifestyle data such as smoking status, exercise and body mass index is of benefit to both the health care providers and the patients.

Effective EPR implementation and networking could eventually save more than $81 billion annually—by improving health care efficiency and safety—and that HIT-enabled prevention and management of chronic disease could eventually double those savings while increasing health and other social benefits.(Hillestad et al., 2005) DiScRi data provides important baseline information for health care quality improvement at local, state and national levels. Including pro-coagulation and pro-inflammatory factors improves overall health care planning as the physician has additional information.(Navab et al., 2006) Similarly, options for viewing retinal images that indicate early signs of diabetes or presence of CVD seen on 12-lead ECG or peripheral pulse wave analysis can be of use to general practitioners. The database allows access to all biochemistry and pathology results as well as the retinal photography, 12-lead ECG and Doppler peripheral vessel blood flow velocity traces. Making the DiScRi EPR the most comprehensive primary health care database available in Australia.

REFERENCES

Bates, D. W., Ebell, m., Gotlieb, E., Zapp, J. & Mullins, H. C. (2003) A Proposal for Electronic Medical Records in U.S. Primary Care. *J Am Med Inform Assoc,* 10**,** 1-10.

Burns, J. R., Bonney, M.-A., Davies, P. G. P. & Harris, M. F. (2004) Utilization of the CARDIAB(R)TM database system to promote quality of care in Australian general practice. *Health Informatics Journal,* 10**,** 177-184.

Grundy, S. M., Pasternak, R., Greenland, P., Smith, S., Jr. & Fuster, V. (1999) Assessment of Cardiovascular Risk by Use of Multiple-Risk-Factor Assessment Equations : A Statement for Healthcare Professionals From the American Heart Association and the American College of Cardiology. *Circulation,* 100**,** 1481-1492.

Hillestad, R., Bigelow, J., Bower, A., Girosi, F., Meili, R., Scoville, R. & Taylor, R. (2005) Can Electronic Medical Record Systems Transform Health Care? Potential Health Benefits, Savings, And Costs. *Health Aff,* 24**,** 1103-1117.

Jelinek, H. F., Wilding, C. & Tinley, P. (2006) An innovative multi-disciplinary diabetes complications screening programme in a rural community: A description and preliminary results of the screening. *Australian Journal of Primary Health,* 12**,** 14-20.

JOSHY, G. & SIMMONS, D. (2006) Diabetes Information Systems: A Rapidly Emerging Support for Diabetes Surveillance and Care. *Diabetes Technology & Therapeutics,* 8**,** 587-597.

Klein Woolthuis, E. P., De Grauw, W. J. C., Van Gerwen, W. H. E. M., Van Den Hoogen, H. J. M., Van de Lisdonk, E. H., Metsemakers, J. F. M. & Van Weel, C. (2007) Identifying people at risk for undiagnosed type 2 diabetes using the GP's electronic medical record. *Fam. Pract.,* 24**,** 230-236.

Michos, E. D., Nasir, K., Braunstein, J B., Rumberger, J A., Budoff, M. J., Post, W. S. & Blumenthal, R. S. (2006) Framingham risk equation underestimates subclinical atherosclerosis risk in asymptomatic women. *Atherosclerosis,* 184**,** 201-206.

Mohiddin, A., Naithani, S., Robotham, D., Ajakaiye, O., Costa, D., Carey, S., Jones, R. H. & Gulliford, M. C. (2006) Sharing specialist skills for diabetes in an inner city: A comparison of two primary care organisations over 4 years. *Journal of Evaluation in Clinical Practice,* 12**,** 583-590.

Navab, M., Anantharamaiah, G. M., Reddy, S. T., Van Lenten, B. J., Ansell, B. J. & Fogelman, A. M. (2006) Mechanisms of disease: proatherogenic HDL - an evolving field. *National Clinical Practice, Endocrinology and Metabolism,* 2**,** 504-511.

Newnham, H., Colagiuri, S., Hepburn, A., Chen X. M. & Colagiuri, R. (2004) National Evidence Based Guidelines for Type 2 Diabetes: Prevention and Detection of Macrovascular Disease. Canberra, NHMRC.

NHMRC (2004) National Health and Medical Research Council. National evidence based guidelines for the management of type 2 diabetes mellitus: prevention and detection of macrovascular disease. Sydney, Australian Centre for Diabetes Strategies.

Penn, D. L., Burns, J. R., Georgiou, A., Davies, P. G. P. & Harris, M. F. (2004) Evolution of a register recall system to enable the delivery of better quality of care in general practice. *Health Informatics Journal,* 10**,** 165-176.

Sheridan, S., Pignone, M. & Mulrow, C. (2003) Framingham-based Tools to Calculate the Global Risk of Coronary Heart Disease. A Systematic Review of Tools for Clinicians. *Journal of General Internal Medicine,* 18**,** 1039-1052.

Woodward, A., Wallymahmed, M., Wilding, J. & Gill, G. (2006) Successful cardiovascular risk reduction in Type 2 diabetes by nurse-led care using an open clinical algorithm. *Diabetic Medicine,* 23**,** 780-787.

E-ORDERING IN THE PHARMACEUTICAL SUPPLY CHAIN
Explaining Standardisation from a Collective Action Perspective

Stefan Schellhammer, Kai Riemer and Stefan Klein
European Research Center for Information Systems (ERCIS)
Department of Information Systems, The University of Münster
Leonardo-Campus 3, Münster, Germany
stefan.schellhammer@ercis.de, kai.riemer@ercis.de, stefan.klein@ercis.de

Keywords: Pharmaceutical Industry, Collective Action, Standardisation, Interorganisational Information System (IOIS).

Abstract: In this paper, we discuss a unique case of industry-wide standardisation, i.e. the proliferation of an electronic ordering protocol across wholesalers and community pharmacies in the Republic of Ireland. The existence of multiple parties involved in the standardisation process and the nature of the standard lead us to study the case from a collective action perspective. In doing so, the emergence and the diffusion of industry-wide standards are being studied as distinct but connected set of dilemmas. The case leads us to theorise that strong industry associations play a significant role in the initiation and success of such standardisation efforts on the industry level. Due to space restrictions this short paper can only provide a snapshot of our entire argument; a long version can be obtained from the authors.

1 INTRODUCTION

The formation of interorganisational information systems (IOIS) has been widely studied in the IS literature. Typically, the purpose of an IOIS lies in supporting, facilitating, or improving inter-organisational (business) transactions, with electronic data interchange as the core building block. In this paper, we specifically focus on vertical information systems (VIS) that promote data exchange and business process coordination between business partners along the supply chain. In doing so, we concentrate on the development of core standards as a prerequisite for the development of interoperable systems among the business partners. The core standards encompass communication protocols, message syntax and semantics, such as product codes. Despite the large body of literature on standardisation and IOIS it still remains unclear why in some industries open standardised IOIS have emerged whereas in others competing systems developed.

IOIS are sometimes regarded as a strategic device to improve customer retention through lock-in, while in other cases IOIS may serve as a means to collectively reduce transaction costs. In this paper, we will concentrate on the second type and briefly juxtapose it with an example of the former type. Our case covers the standardisation process in the creation of a universal electronic ordering system in the Irish pharmaceutical distribution, which took place during the 1980s. In contrast to the standardised Irish solution, in the British pharmaceutical distribution industry several competing electronic ordering system have emerged.

Our data shows that the resulting differences cannot be explained by environmental factors like regulation or by the needs of their users; most of these factors are strikingly similar. The question arises what triggered the set up of the Irish system in this unique way. We will argue that an industry association has played a significant role in this process, and that the role and importance of associations for the emergence of IOIS has not been studied sufficiently (Damsgaard & Lyytinen, 2001).

We are facing two challenges: (1) How can we explain that the initiators or sponsors of electronic ordering solutions (in our case the wholesalers) partially suspended their competition and agreed on standards to be used for electronic ordering? (2) How can we explain the adoption and diffusion across a heterogeneous and fragmented group of business partners, i.e. the pharmacies. We will approach this effort from a collective action theory perspective. In our analysis we draw on earlier work conducted in this field (Markus *et al.*, 2006).

In the next section we present the Irish case. After this we introduce collective action theory before we analyse and discuss the case.

2 IRISH PHARMACEUTICAL DISTRIBUTION INDUSTRY

In this section we document how EDI standards emerged in the Irish pharmaceutical distribution between wholesalers and community pharmacies. We begin by introducing our method and give a brief overview of the pharmaceutical industry in Ireland. This is followed by a description of the emergence of the standard itself and its subsequent use by the actors.

2.1 Method

A case study design has been chosen to conduct the research, because of the complexity of the research question and its focus on a rich real-life context (Yin, 2003). Four semi-structured interviews have been conducted, two with a representative of one of the Irish pharmaceutical wholesalers, one with leading managers at the Irish body of community pharmacists (IPU) and one with a manager of a large software system vendor. All interviews were tape recorded, transcribed, coded and analysed. The data was evaluated independently by two researchers. Furthermore, several other data sources were used for the study – among these are web sites, standards documents, systems documentations etc.

2.2 Market Structure

The task of the pharmaceutical *wholesalers* is to provide a national wholesale service for pharmaceuticals for community pharmacies and hospitals. The relevant market is demarcated by the national borders. As a result of a consolidation process over the past 10-15 years, the market in the Republic of Ireland (R.I.) is nearly evenly divided between three wholesalers. All Irish wholesalers operate as nationwide full-line suppliers.

On the customer side more than 1400 *community pharmacies* exist in Ireland. Irish pharmacists are represented by a professional body, the *Irish Pharmaceutical Union (IPU),* which represents 90% of all Irish pharmacists. The mission of the IPU is to promote the professional and economic interest of its members. This incorporates conducting negotiations on behalf of the members and the development and maintenance of a "constructive dialogue with government, agencies and other groups in relation to matters of mutual interest." (IPU, 2006)

Three *software vendors* serve the pharmacy market. Their product, the "patient medication record" (PMR) incorporates the EDI standards as basis for the ordering module. Around 1300 pharmacies work with computer systems (typically with electronic point of sales systems (EPOS) and dispensary software).

The *wholesale prices for drugs* are fixed as is the margin for wholesalers. However, the wholesalers de facto pass on a significant part of their margin to the pharmacies via rebates, bonus schemes and other price incentives (Fingleton *et al.*, 2002). These prescription drugs are paid by the patients or they are reimbursed to the pharmacies through a variety of state-administered schemes (GMS, DPS, LTI).

As a result of these regulations the use of *pricing policies* by wholesalers and pharmacies is very restricted. As wholesalers typically offer volume discounts, pharmacies generally use one wholesaler as the primary supplier and a second one to split purchases and as a fall-back when supply of a particular product cannot be obtained (Fingleton et al., 2002). While the wholesalers compete for becoming a pharmacy's preferred supplier, the pharmacies use their bargaining power as a result of low switching costs: The *possibility to switch between wholesalers* alleviates the problem of stock shortage on the wholesalers' side, because they risk loosing pharmacies in case of repeated stock shortages. *Swift delivery* is a crucial element of the pharmaceutical supply chain. Consequently, most pharmacies operate on very small stock. This is made possible by extensive logistics operations. Pharmacies can rely on very short delivery times and deliveries two times a day by each of the wholesalers. Wholesalers will ship on the same day all orders that are received by the cut-off time late in the morning.

2.3 Emergence of the Standard

In 1984 United Drug (UD), one of the wholesalers, studied the emergence of electronic ordering systems (McKesson, see (Johnston & Vitale, 1988)) and intended to adapt one of the U.S. solutions for the Irish market. The Irish pharmacies, in dispensing medicines, only kept handwritten records at the time.

Because pharmacy market regulation varies significantly across countries, UD decided to develop a new solution from scratch. In doing so, UD played with the idea of developing a proprietary ordering system; the strategic rationale being to lock-in

pharmacies and subsequently to increase market share. The idea was to take the UK market as a blueprint where American Hospital Supply had successfully established ordering software to lock-in hospitals.

While the wholesalers claim that they ultimately realised the shortcomings of a proprietary solution it was the Irish Pharmaceutical Union (IPU) who strongly engaged in the process and came out with an open solution. While the IPU regarded the move from placing orders over the phone to submitting electronic order files as a clear administrative advantage for the pharmacies (faster, less errors etc.), key actors at the IPU emphasised the benefits of a standardised solution over a number of proprietary solutions for the pharmacies.

As a result, the IPU facilitated negotiations among the wholesalers to develop both a standardised order data transfer protocol and a common numbering system (based on EAN). While wholesalers in the UK were introducing product numbering based on the PIP code, IPU had been negotiating with EAN (at that stage it was EAN UK, today it is GS1 Ireland) to use its numbering scheme. The idea was to administer a central product identification number that does not differ between wholesalers, as was (and is) the case in the UK (Chemist&Druggist, 2007). In order to facilitate the introduction of EAN numbering codes, the IPU itself applied for and was granted manufacturer status in order to be able to assign EAN numbers to pharmaceuticals.

In parallel, the IPU facilitated negotiations among the wholesalers and system vendors. Consensus on a common protocol was achieved after about 6 months. At the time, half a dozen system vendors developed pharmacy solutions and all of them included the electronic ordering standards into their packages. Conceptually, this led to the proliferation of different IOIS, all of which are using the same standardised EDI components (see Figure 1).

2.4 System Usage and Maintenance

The diffusion of electronic ordering happened gradually over a 10 year period; today, all pharmacies are able to order electronically and the transactions between wholesalers and pharmacies are still based on the same system. Surprisingly, even the modem-based communication protocol has survived virtually unchanged. Electronic ordering (eOrdering) accounts for 90 percent of all orders.

The wholesalers' systems accept incoming orders (wants lists) via a modem connection and send back an order confirmation. If all items on the "wants list" can be supplied no response is sent. In any other case the systems send back a list with unavailable items. Furthermore, bonus items are reported back to the pharmacist in this way. The pharmacy can then immediately turn the list of out of stock items into an order with one of the other wholesalers.

The IPU-product file is a complete list of all products available in an Irish pharmacy. It identifies each item by an IPU-number. Furthermore, the file also contains characteristics of the product (package size, ingredients and toxic class), consultation advices, prices and additional codes for reimbursement. For maintenance and administration of the file, the IPU operates a dedicated department. The IPU distributes the product file on a monthly basis to all pharmacies, wholesalers and system vendors.

The order protocol uses the IPU product file for product identification and the IPU communication protocol called IPUCOMMS v 2.4a for the modem dial-up link from the pharmacy to the wholesalers.

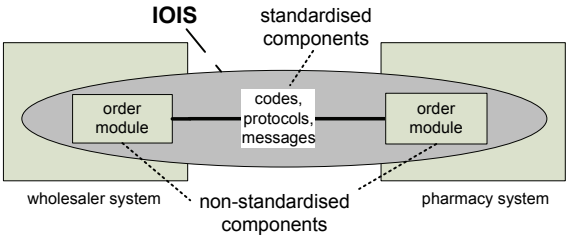


Figure 1: Components of the IOIS.

3 COLLECTIVE ACTION

The case presented above shows the development of common EDI standards, which are the cornerstone of the electronic ordering system that is still used in Ireland today. We argue that the presented phenomenon of joint standard development can be interpreted as a case of collective action.

The Theory of Collective Action deals with the provision of public goods. It explores the market failures where individual rationality and self-interest do not lead to an efficient provision of public or collective goods. Please note that, in our case, we are looking at standards as an instance of non-pure public goods, i.e. club goods that belong to a sub group of the market (they are excludable (Kindleberger, 1983)), but share similar characteristics.

Markus et al. have studied the consequences that collective good characteristics have on standardisation. (Markus et al., 2006) They differentiate the

development and the diffusion of standards as two distinct dilemmas. They come to the conclusion that for a successful standardisation both need to be solved at the same time.

The *dilemma of the development of a standard* refers to the different areas of conflict that exist when a standard is developed in a consortium: Conflicts of interest refer to the heterogeneity of interests among the actors; conflicts of alignment refer to the cost of aligning internal systems to the new standard; and conflicts of appropriation result from different cost/ benefit structures of the actors (Müller-Tengelmann, 1995). Mechanisms have to be found to solve these conflicts to ensure successful standards development.

The *dilemma of the diffusion of standards* refers to the incentive to delay the adoption of a standard until a critical mass has been reached. The diffusion of public goods (e.g. communication standards) is often hampered by what is called the penguin effect (Farrell & Saloner, 1987): Early adopters cannot capitalise fully on the standard as long as no other actor adopts the standard (negative network effect).

The interconnection between both dilemmas has been described by Markus et al.: "…standards content can be seen both as an outcome of the mechanisms employed by a VIS standards-setting consortium to resolve collective action dilemmas and as an input to diffusion on the VIS standards developed."(Markus et al., 2006)

The successful development of a standard is the conditio sine qua non for its diffusion. However, it does not guarantee the success of the standard in the market. To the contrary, strategies to solve the development dilemma can turn out to be counterproductive for its diffusion. Hence, the dual dilemma situation can only be solved if both dilemmas are targeted simultaneously.

4 CASE ANALYSIS

In this section we analyse the driving factors behind the standardisation process in the Irish case using collective action as the conceptual framework.

4.1 Standards Development

While different actors claim ownership of the idea of a standardised product code, the IPU certainly has been one of the most vocal advocates of standardised solutions at a time when some of the wholesalers were still considering proprietary solutions. Their rationale was threefold: a single standard in the Irish market would facilitate the existing practice of order splitting and thus maintain the lever by which the pharmacies executed their (small) power vis-à-vis the much bigger wholesalers. Secondly, standards-based electronic ordering would be the most efficient way of ordering and would help to reduce transaction costs. Thirdly, the Irish market is too small to justify competing solutions.

Moreover, the IPU clearly played a crucial role in facilitating and moderating the negotiations between the wholesalers. The IPU provided a neutral venue and drove the negotiations without being perceived as partisan. In effect they facilitated joint action of the wholesalers. The negotiations took about six months which were regarded as efficient by the participants.

In doing so, the IPU played a triple role: Next to the *moderator or broker role*, the IPU represented the pharmacists and their interests. Given that they represented the overwhelming majority of pharmacists, they had a strong mandate to articulate their constituents' interests. Through this *representative participation*, they not only pushed hard for a standardised solution but also shaped the design of the standards. Moreover, the IPU provided some initial assurance that the solution would be adopted by the pharmacies, the prospective adoptors: the IPU de facto overcame the fragmentation of the pharmacies and ensured that the new system would perpetuate and indeed facilitate the established practices of order splitting. Thirdly, the IPU became the *secretary of the standardisation process*; in particular they agreed to take the role of developing and maintaining the product codes (IPU product file). Thereby they ensured the sustainability of the chosen solution and reinforced their role in the market by securing an additional revenue stream based on the license fees for the product file.

All in all the IPU managed to contain the conflicts of interest among the wholesalers (horizontal conflicts of interest) as well as potential conflicts of interest between the wholesalers and the pharmacies (vertical conflicts of interest).

The alignments required of the internal systems, as another potential area of conflict, could be kept to a minimum by developing a product code that conforms to the EAN-13 standard that was already being processed by the wholesalers. Thus, no major alignments were required. With respect to the ordering protocol, the negotiating process has been described as very cooperative in the sense that different needs of the participants were accommodated. Under the prime aim to hammer out a common protocol, the wholesalers tried to make sure that it

would dove-tail with their systems. Furthermore, the scope of the solution was kept lean so that alignments were kept to a minimum.

Conflicts of appropriation are the third area of potential conflict. Our data reveals no indication of this type of conflict. Ordering protocol and product file are administered and maintained by the IPU. The IPU refinances these activities through licensing fees from wholesalers and manufactures. Our data does not indicate that this division of tasks has ever been challenged by the actors.

The case shows a successful development of a standard. This is especially interesting as in other countries with similar characteristics like the UK (but also Australia) proprietary solutions exist. Good reasons can be found for either option: competitive advantages and customer lock-in on the one side, efficiency and swift diffusion on the other.

However, in the Irish case it was the IPU who facilitated the development of a joint solution. As the wholesalers are the main beneficiaries from electronic ordering, the IPU was in a position to articulate requirements on behalf of the pharmacies while providing assurances of the adoption and diffusion of the solution. Today, the product numbering scheme can be seen as a new power basis for the IPU. The IPU succeeded to preserve that power basis against the PIP-code solution of the wholesalers. Thereby the IPU was able not only to satisfy needs of its members but also to bring itself in a better position for future negotiations. In the end, the wholesalers made themselves dependent on the IPU and its product code.

4.2 Standards Diffusion

While the literature reports cases where the parties involved in the standard development later blocked or delayed their diffusion, the Irish pharmaceutical wholesalers embraced and supported the new standards. All wholesalers have integrated the new ordering standards into their own systems. The same is true for the system vendors supplying the pharmacies with the software.

While the IPU encouraged its members to install and use the new electronic ordering facility, it is reported that in the beginning pharmacists have been reluctant to use the new technology; they saw the advantages of the new system on the wholesalers' side. The wholesalers initially responded with discounts on electronic orders. Today, about ninety percent of all orders reach wholesalers electronically.

The high implementation ratio of wholesalers can be explained by the low costs of alignment and a lack of feasible alternatives due to the bargaining power of the pharmacists represented by the IPU. No wholesaler tried to impede the development process openly. And any attempt to promote a proprietary system by an individual wholesaler afterwards would not have been tolerated by the pharmacies, but would likely have encouraged them to switch to the other wholesalers. Furthermore, all system vendors took part in the negotiations. A multiplicity of proprietary standards would not have been in their interest, because pharmacists would have pressured them to build software incorporating all different standards.

Several factors can be identified that favoured the industry-wide standard and ultimately triggered its success: First, the IPU has a strong standing towards the wholesalers. The wholesalers were very aware that anything that would run against the interest of the pharmacists would face strong and painful opposition by the IPU. Furthermore, the IPU was not taken by surprise when the wholesalers started to develop electronic ordering solutions. Rather, the IPU was aware and attentive towards these new technologies and their implications for its members. This set the IPU in a position to intervene at an early stage, where the different stakeholders had not yet firmly committed themselves nor invested into a particular technology. Timing was critical and the IPU clearly used it to its advantage.

Another factor that facilitated the industry-wide standardisation can be seen in the relatively small group of wholesalers operating in a small and confined market. Regulated prices and margins pose a high pressure on wholesalers to optimise their processes and save costs in warehousing, delivery and order processing. Therefore a high incentive existed to get such a system working and thereby streamlining the order process.

5 CONCLUSIONS

While we are interested in the ordering system as a VIS or IOIS, our analysis has focused on the underlying standards as core prerequisites of interorganisational solutions. The Irish case reinforces the notion that standardisation processes for VIS are indeed precarious. Proprietary solutions have been considered by the initiators and have been chosen in other countries with plausible strategic motives.

Our data supports the notion of a dual dilemma of standardisation, which needs to be addressed si-

multaneously. The dilemma implicitly also addresses the need to overcome horizontal competition (among wholesalers) and vertical conflicts of interest (among wholesalers and pharmacies).

With respect to our research questions we have found theoretical explanations for the unique case situation. The single most important part of our explanation rests on the role, power and reputation of the IPU. The IPU facilitated to initiate and moderate collective action initially among the wholesalers and subsequently with an increasing mandate and role for themselves as party to the negotiations. By representing the overwhelming majority of the pharmacists they overcame the fragmentation of the market and changed the power dynamics in the negotiations. Moreover, they provided assurances with respect to the adoption of the standards. The IPU established themselves successfully as standards keeper. The availability of widely accepted standards like EAN product coding schemes clearly helped to build credibility and to enhance the acceptance of the chosen solution.

The wholesalers as initiators of ordering systems agreed on common standards because they saw (or were alerted to) the obvious interests of the pharmacies as potential adopters.

However, the constellations of actors, historical and regulatory environment has been quite unique. While we have found theoretical explanations for the outcome ex post, similar developments are still far from predictable. The notion of causality remains contested: too many contingencies and considerations are at play, which could have lead to another outcome. Hence we have tried to establish plausible reasons.

The specific actor constellation, in particular the multiple roles which the IPU played successfully, explain the achieved consensus. However, there were a number of facilitating contingencies, which have not determined the outcome but help to explain it: The historical coincidence of standard development in a technological "virgin market", where the business partners had not yet invested in their own systems, convinced all parties – including the software vendors – to pursue the chosen standards. The area of consensus building (product code, order message, communication protocol) facilitated the consensus building. The standards guaranteed interoperability between the applications yet in a model of loose coupling.

In the end, the strong role(s) of the IPU, in combination with the economic benefits (and incentives) of electronic ordering and the existence of one uncontested standard, lead to wide adoption and a sustainable solution.

ACKNOWLEDGEMENTS

The study described in this paper is part of a research project (no. 1328/2-2) funded by a research grant of the DFG (German National Science Foundation) and is concerned with the question of how institutional and national factors influence structures and evolutionary paths of IOISs.

REFERENCES

Chemist&Druggist. (2007). The PIP code in the pharmaceutical supply chain. Retrieved 16.07., 2007, from http://www.dotpharmacy.co.uk/pip.html

Damsgaard, J., & Lyytinen, K. (2001). The Role of Intermediating Institutions in the Diffusion of Electronic Data Interchange (EDI): How Industry Associations Intervened in Denmark, Finland, and Hong Kong. *The Information Society, 17*(3), 195-210.

Farrell, J., & Saloner, G. (1987). Competition, Compatibility and Standards: The Economics of Horses, Penguins and Lemmings. In H. L. Gabel (Ed.), *Product Standardization and Competitive Strategy* (11 ed., pp. 1-21). Amsterdam: Elsevier.

Fingleton, J., Purcell, D., & Goggin, I. (2002). Report of Investigation (pp. 1-64): Competition Authority.

IPU. (2006). About the IPU. Retrieved 16.07, 2007, from http://www.ipu.ie/index.php?option=com_content&task=view&id=17&Itemid=42

Johnston, H. R., & Vitale, M. R. (1988). Creating Competitive Advantage with Interorganizational Information Systems. *MIS Quarterly, 12*(2), 153-165.

Kindleberger, C. P. (1983). Standards as Public, Collective and Private Goods. *Kyklos, 36*, 377-396.

Markus, M. L., Steinfield, C. W., Wigand, R. T., & Minton, G. (2006). Industry-wide IS Standardization as Collective Action: The Case of the US Residential Mortgage Industry. *Management Information Systems Quarterly, 30*(Special Issue), 439-465.

Müller-Tengelmann, H. (1995). *Kollektive Investitionsstrategien - Der elektronische Datenaustausch als überbetriebliche Infrastruktur*. Frankfurt am Main; Berlin; Bern; New York; Paris; Wien: Lang.

Yin, R. K. (2003). *Case Study Research: design and methods*: Sage Publications Inc.

ICT ARCHITECTURE FOR A COMMUNITY MEDICINE NURSE PROJECT

Thomas Karopka, Ilvio Bruder
IT Science Center Rügen gGmbH, Circus 14, 18581 Putbus, Germany
karopka @it-science-center.de, bruder @it-science-center.de

Neeltje van den Berg, Wolfgang Hoffmann
Section Epidemiology of Health Care and Community Health, University of Greifswald, Greifswald, Germany
neeltje.vandenberg@uni-greifswald.de, wolfgang.hoffmann@uni-greifswald.de

Andreas Heuer
Institute of Computer Science, University of Rostock, Rostock, Germany
heuer@informatik.uni-rostock.de

Keywords: Telemedicine, Tele-homecare, Community medicine nurse, ICT.

Abstract: In the Federal State of Mecklenburg-West Pomerania 35-40% of the general practitioners (GPs) will retire within the next 5-7 years. In rural regions, it is difficult to find successors for the vacant practices. Thus a problem of supplying primary health care to the elderly population in rural regions is foreseeable. An efficient way to lower the workload for the remaining GPs is the implementation of a special trained community medicine nurse (cm-nurse). The cm-nurses are supported by telemedical devices, a video conference system and a mobile data management system. In this paper we report on the information and communication technology (ICT) architecture of the project.

1 INTRODUCTION

Society's demographic change and structural changes in the German health-care system are nationwide challenges. In the Federal State of Mecklenburg-West Pomerania, demographic changes are developing more dynamically than in many other regions in Germany. At the end of 2005, 25.4% of the population was more than 60 years old - a percentage which will climb to 34.8% by the year 2020. One of the consequences of an ageing population is the increase of age-specific diseases.
On the other hand, 35-40% of the general practitioners (GPs) will retire within the next 5-7 years.
In rural regions, it is difficult to find successors for the vacant practices. Thus a problem of supplying primary health care to the elderly population in rural regions is foreseeable. The Institute of Community Medicine of the University of Greifswald and the Institute of Computer Science of the University of Rostock have started an interdisciplinary research project between the faculties of medicine, health economics, pharmaceutics and informatics to develop a model project for a cm-nurse that supports GPs in the primary health care for elderly patients in rural areas (Terschüren, 2007). One of the main goals of the project is to develop a model that disburdens the GPs in rural areas in the task of serving an ever increasing area and therefore spending a lot of time travelling from the patients for home visits. The idea is that a part of the home visits can be carried out by a special trained cm-nurse, supported by telecare equipment. Modern mobile technologies, a carefully designed information flow between the cm-nurse, GP and the participating pharmacists as well as a central data store for epidemiological and health economic evaluation are the prerequisites for the development of a new model that is unprecedented in the German health care system.
A key role for the success of a project of this kind plays the underlying ICT infrastructure. While the different devices and techniques needed already exist,

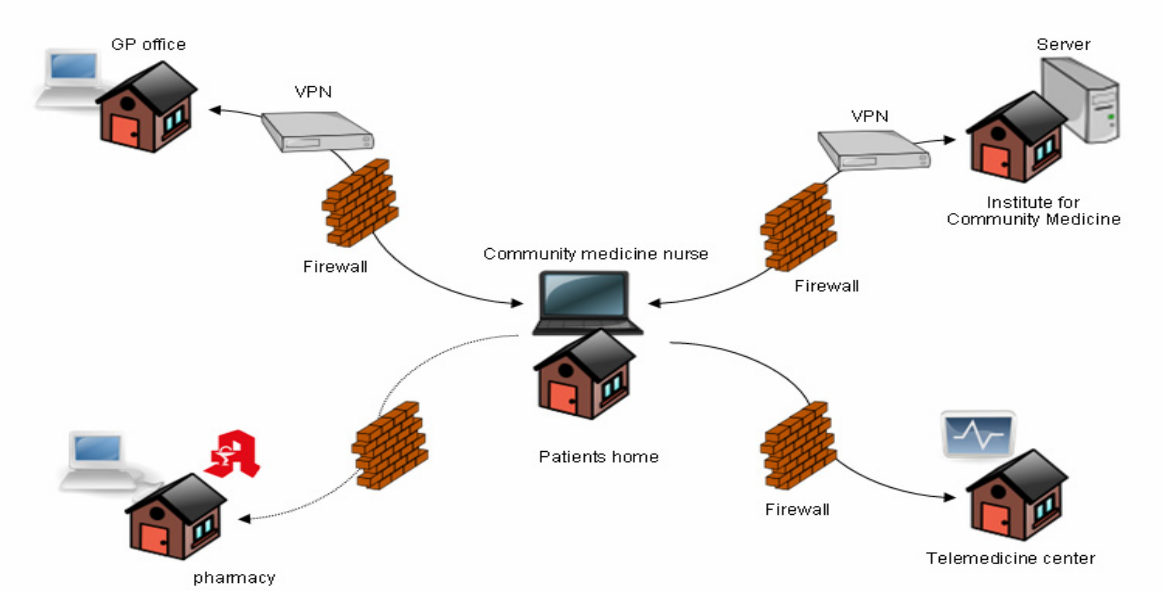


Figure 1: System Architecture.

the integration of these into a working system is still a challenge. In this paper we focus on the technological aspects and challenges for the ICT infrastructure needed to allow the project to work.

2 DESIGN OF THE SYSTEM

Several requirements had to be fulfilled and influenced the design of the system. The main requirement was that the system should work in rural areas with no guarantee of internet access. Further, the system should be open for different clients and operating systems. Another requirement was the usability of the system. The software should be designed in a way that reflects the workflow that the nurse is accustomed to.

Figure 1 gives an overview of the system. The central part is a tablet-PC that is used by the cm-nurse during the home visits. The Institute for Community Medicine has developed standardised questionnaires to support the cm-nurse in the daily work and to collect data for research. At the end of a working day the cm-nurse transmits the data to the central database. To support the research in the project, a central database with pseudonymized patient data was installed in the Institute of Community Medicine. The database consists of a MySQL database currently running on a Windows 2003 server behind the university firewall. The cm-nurses use a VPN client to log into the university network to access this central database. Several other interfaces exist in the system. To support the drug anamnesis module data is exchanged between cooperating pharmacies the GP and the cm-nurse. The data that the nurse collects can be sent to the electronic patient record (EPR) system in the GP office. Several telecare devices are also used in the system. This interface is currently not integrated into the system and is not considered here.

3 IMPLEMENTATION

Whenever possible we used open source software for the system. The system was implemented as a client server system. We used the model-view-controller (MVC) design pattern for the development of the system. This allows for easy implementation of different clients. Figure 2 illustrates the software design. Currently the only clients used are tablet PCs. But it is also possible to use PDAs or UMPCs. Apache Struts is used as the framework. Tomcat 5.5 is used as servlet container. The recorded data is stored in a local object database (DB4O) on the tablet PC. At the end of a working day the cm-nurse transmits the data to the server in the Institute for Community Medicine using a VPN connection. On the server the data is inserted in a MySQL database. This is done using Hibernate for the

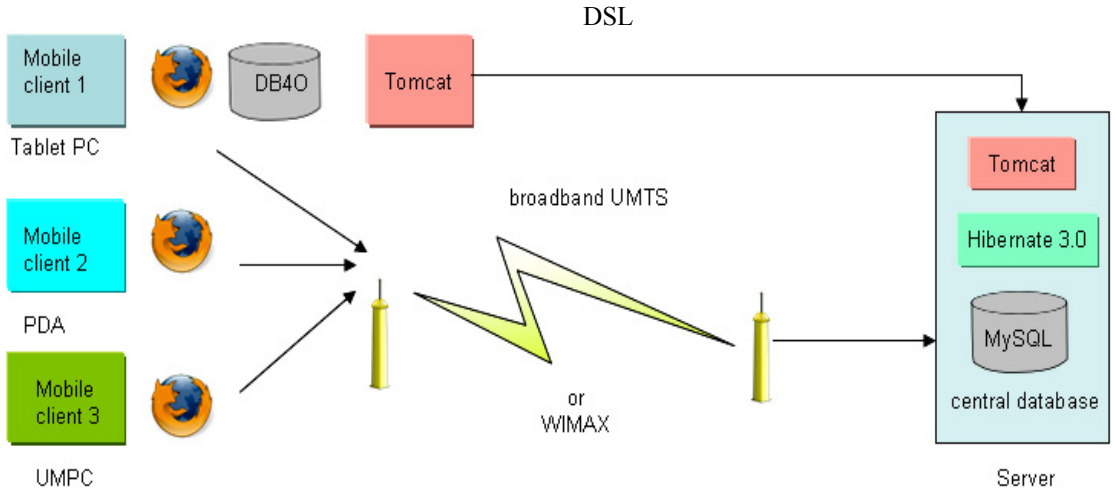


Figure 2: Software Design.

object-relational mapping. The software was designed in a way that it is possible to add other mobile clients in a later phase of the project.

4 DISCUSSION

The constraint with the highest impact on system design was the requirement for the system to work in rural areas. This is a contradiction to the need of a broadband wireless infrastructure. In the first pilot project we experimented with WIMAX but it was not possible to serve the area needed. An alternative is GSM/GPRS or broadband UMTS. Both system were not available in all of our test regions. We also noticed problems with unstable connections. Therefore we decided to implement a system that allows the nurse to work offline. The nurse selects the patients she wants to visit. All the data is copied to a local database on the tablet-PC. Once the nurse is back in her office, she connects the tablet-PC to the internet and transmits the data to the central server.

Another process was the selection of the client. We proposed three clients: A tablet-PC, a PDA and a UMPC. PDAs are already in use in different clinical settings and also in home care (Hsu, 2007; Alonso, 2004). Although the PDA is small, light and has a long battery runtime, we opted against the PDA because the display is too small for our application. The modules are designed to guide the cm-nurse through her visit with standardised questionnaires. On a PDA the handling was judged to be too complicated. The UMPC was ruled out due to the low battery runtime. We finally choose the tablet PC as a compromise. It has a good battery runtime, allows a good visualisation of the questionnaires and is easy to handle with the pen.

Table 1 shows the kind of data and the number of measurements in the monitoring module in one of the projects.

Table 1: Type and frequency of data in the monitoring module.

Module	**# of Measurements**
Monitoring	1822
Blood pressure	1696
Pulse	1241
Drink monitoring	1142
Blood sugar	839
Weight	486
Blood sample	364
Nutrition advice	382
Medical advice	273
Conversation	126
Vaccination	125
Wound examination	112
Dressing	101

The decision to use the Apache Struts Framework was partly due to time constraints in the project. We

only had one month to develop the core of the system. The Apache Struts framework with the validation support and the supplied libraries was an ideal starting point for a prototype. Later on it was very easy to add new modules to the software and it was decided to keep the web application instead of developing a native client. A disadvantage of the servlet technology is the response time that the servlet container needs to build the page and the stateless interface. The advantage, however, is the easy to use interface and the maintainability of the software. Currently the software is used in four different field trials by 27 cm-nurses.

To the authors knowledge the model of a community medicine nurse supported by ICT as described here is new. However, the participation of nurses in primary care or the support of GPs by nurses is already considered in several countries. In the systematic review of (Laurant, 2005) the authors come to the conclusion that appropriately trained nurses can produce as high quality care as primary care doctors and achieve as good health outcomes for patients. The report of (Bourgueil, 2005) compares the situation in six European countries as well as in Quebec and Ontario. However, there is few information of the ICT used by nurses to communicate with the GP. Most of the published literature is concerned with telenursing or telehomecare. Telenursing refers to the use of telecommunications technology in nursing where the nurse either does a "virtual visit" using a video conference system or a telephone to provide home care services. Studies in Europe suggest that a large number of patients could benefit from in-home telecommunication services (Valero, 1999). Telehomecare means the delivery of health services over distances into the home care setting where home health nurses use technology to provide services in the home which enhance the efficiency and the quality of care (Milholland, 1995). In these use cases ICT is used to support a home health nurse in providing her service, usually over a distance. Although we also use telemedical devices the primary use case of our software is the standardised communication of a cm-nurse with the GP. The cm-nurse works in delegation of the GP i.e. a task formerly carried out by the GP is now transferred to the cm-nurse. However, the GP needs to have the overview of the health state of the patient and therefore needs the measurements as well as the standardised questionnaires to get an impression of the patients' current health state. Table 2 gives an overview of the currently implemented modules.

Table 2: Overview of implemented modules.

Module	Description
First Interview	General questions. Carried out during first patient contact
Standard monitoring	Module that is carried out during every home visit
Training for telemedical devices	Used to document the training in the use of the telemedical devices
Fall prevention	Standardised questionnaire for the detection of fall risks in the domesticity of the patient
Drug anamnesis	Registration and check of all drugs in the domesticity of the patient (including interaction check in cooperation with local pharmacist)
Geriatric assessment	Test for signs of possible dementia: clock drawing test (Shulman, 1986), DemTect (Kessler, 2000)
Palliative care	Management of pain and provision of psychological and social support
SF-12	Short form of the SF-36 health survey for assessment of the health related quality of life of patients (Bullinger, 1995)

A project that uses very similar technology is from the Luleå University of Technology in Sweden (Andersson, 2007). The project is called SARAH and is run by the Center for Distance spanning healthcare at Luleå University of Technology, Norrbotton County council and the municipalities in Luleå and Boden. In this project a district nurse is supplied with a field rucksack with a laptop facilitating mobile access to the electronic patient record and videoconferencing between the GP and the district nurse, an electronic stethoscope, a digital camera and other telecare equipment.

For future work an interface for the cooperating pharmacies is planned where the pharmacist can

access the questionnaire for the drug anamnesis module.

5 CONCLUSIONS

An ageing population, fewer GPs in rural areas and high costs in the healthcare system challenge the development of new models in home healthcare. ICT offers numerous potential benefits in terms of improvements for patients, professionals and cost savings in the healthcare system. We have implemented a complex ICT infrastructure to support a community medicine nurse that carries out home visits in delegation of a GP. The ICT infrastructure supports the cm-nurses in their daily work as well as providing data for the health-economical and epidemiological evaluation of the project.

ACKNOWLEDGEMENTS

The authors would like to thank Andreas Holtz who designed and implemented the original system.

REFERENCES

Alonso, A. 2004. A new model for home care for COPD. *Stud.Health Technol.Inform.,* 103, 368-373.

Andersson, S., Andersson, A., & Sävenstedt, S., 2007. SARAH- Greater Trust in Home-Based Care. *Proceedings of the Tromsø Telemedicine and eHealth Conference 2007*.

Bourgueil, Y., Marek, A., & Mousquès, J. 2005. The participation of nurses in primary care in six European countries, Ontario and Quebec. *Health economics letter – Issues in health economics*.

Bullinger, M., Kirchberger, I., & Ware, J. 1995. Der deutsche SF-36 Health Survey. Übersetzung und psychometrische Testung eines krankheitsübergreifenden Instruments zur Erfassung der gesundheitsbezogenen Lebensqualität. *Zeitschrift für Gesundheitswissenschaften* 3, 21-36.

Hsu, C. L., Kuo, Y.Y., & Chang, P. 2007, Using PDA to Transform the Long MDS-HC Evaluation Form into a Favored System.. *Stud.Health Technol.Inform.,* 129, 1463-1464.

Kessler, J., Calabrese, P., Kalbe, E., & Berger, F., 2000. DemTect: A new screening method to support diagnosis of dementia. *Psycho,* 26, 343-347.

Laurant, M., Reeves, D., Hermens, R., Braspenning, J., Grol, R., & Sibbald, B. 2005. Substitution of doctors by nurses in primary care. *Cochrane Database Syst. Rev.* 2, CD001271.

Milholland, D., K., 1995. Telehealth, telenursing, telewhat? *Am. Nurse*, 27(6), 13.

Shulman, K.I., Shedletsky, R., & Silver, I., 1986. The challenge of time: clock drawing and cognitive function in the elderly. *Int J Geriatr Psychiatry*, 1, 135–140.

Terschüren, C., Fendrich, K., van den Berg, N., & Hoffmann, W., 2007. Implementing telemonitoring in the daily routine of a GP practice in a rural setting in northern Germany. *Journal of Telemedicine and Telecare*, 13(4), 197-201.

Valero, M.A., Arredondo, M.T., del Nogal, F., Rodriguez, J.M., & Torres, D. 1999. Using Cable Television Networks for Interactive Home Telemedicine Services. *Journal of Telecare.* 5(1), 91-92.

A DVB-T BASED SYSTEM FOR THE DIFFUSION OF TELE-HOME CARE PRACTICE

Gianmarco Angius, Danilo Pani, Luigi Raffo, Stefano Seruis
DIEE - Dept. of Electrical and Electronic Engineering, University of Cagliari, Piazza d'Armi, 09123 Cagliari, Italy
g.angius@diee.unica.it, pani@diee.unica.it, luigi@diee.unica.it

Paolo Randaccio
Dept. of Physics, University of Cagliari, Cittadella Universitaria, Monserrato, Italy
paolo.randaccio@ca.infn.it

Keywords: Telemedicine, tele-home care, remote monitoring, ECG, DVB-T, MHP.

Abstract: Typical telemedicine systems are usually PC based and in some cases they use expensive custom devices to satisfy the system requirements. For tele-home care uses by elderly or untrained people, these type of solutions are impracticable. As an example of user-friendly tele-home care system, in this paper we present the first tele-home care application of DVB-T technology over standard home entertainment equipments and a prototypal low cost microcontroller-based acquisition unit for 1-lead ECG. The usability and low cost of the system show the potentiality of the approach.

1 INTRODUCTION

In order to reduce the costs for both the patients and the public administration, tele-home care systems are today often used for clinical practice. They permit to avoid the overcrowding of ambulatories for simple routine examinations and, in the case of chronic patients that must be frequently monitored, to avoid the costs and the waste of time of an ambulatory examination for simple physiological measurements that they could easily perform at home, sending the recorded data to the care staff in the appropriate department. Howewer, these telemedicine systems are usually PC based or use complex hardware and software products. Even if such technologies are quite popular today, they are frequently used only by relatively young people, which are not the primary tele-home care target.

The Digital Video Broadcast Terrestrial (DVB-T) technology, under experimentation in Europe, will be the only terrestrial television system after 2012. By the use of a cheap set-top box, it is now possible to adapt the new transmission tecnology to the old TV equipments. The set-top box is actually a simple computer, normally embedding a RS-232 compliant port connector, a smart card reader and a modem for the return channel link. In principle, it is then possible to define tele-home care systems based on such technology where the user interface is represented by the TV screen (output) and the remote control (input). This paper presents an example of a such system based on DVB-T technology with a cheap microcontroller-based acquisition unit and a very user-friendly interface well exploitable even by elderly people.

The remainder of this paper is organized as follows. In Section 2 a brief analysis about related works is presented. In Section 3 the whole proposed system is introduced; Section 4 deals with some details on the prototypal 1-lead ECG Base Station, whereas the application on the set-top box is presented in Section 5. Section 6 concludes this work.

2 OTHER TELE-HOME CARE PROJECTS IN LITERATURE

The primary activity of telemedicine systems was the transmission of diagnostic medical images using television for medical consultation from physician to physician in remote places. Recently, telemedicine systems through the Internet via satellite have become a reality by means of high throughput mixed satellite-web communication channels. This Section presents some home-oriented telemedicine projects making use of DVB and Internet, without the claim to be exhaustive.

The Interactive Satellite Multimedia Information System (ISIS) project (Pierucci and DelRe, 2000), realized a telemedicine system based on satellite com-

munication for furnishing interactive services for residential users together with the traditional TV distribution, porting tipical applications developed for terrestrial network (Internet) to the satellite digital video broadcasting technology (DVB-S). By a dualband terminal, connected in uplink via Eutelsat satellite and in downlink via the Italsat satellite, the ISIS system provides a small and low-cost transceiver for sites not connected to the Internet or connected only in dial-up low-speed mode. Through his Java application for the DVB-S transceiver, called Medical Environment for Diagnostic Images (MEDI), the project can manage a medical image database for remote expert consultation and so it demonstrated the feasibility of satellite-based interactive multimedia services for telemedicine purposes.

In 2001, a multipurpose health care telemedicine system with a base unit and a mobile unit was developed in Greece too (Kyriacou et al., 2001). At patient's home, the mobile unit allows the transmission of vital bio-signals and static images of the patient to the base unit at the physician site (office or hospital). The mobile unit device is compliant with some of the commercial main vital signs monitor and it is able to transmit ECGs, non-invasive blood pressure (NIBP), body temperature, percentage of arterial oxygen saturation (SpO_2) and heart rate. Based on the TCP/IP protocol, the communication between the two parts ensures safe data transmission and the possibility to use different telecommunication means (GSM or satellite links).

Launched in January 2001, the Universal Remote Signal Acquisition For hEalth (U-R-Safe) tele-home care project (Mailhes et al., 2003), created a mobile telemedicine system for home monitoring to be used by elderly peaple and disabled patients in Europe. Via short range Wireless Personal Area Network (WPAN), wearable ECG and SpO_2 sensors are connected to a portable electronic device, able to send the recorded data to a remote central through the TCP/IP protocol and the wireless public network (GPRS, UMTS and GEO satellites). The portable unit is also capable of sending an alarm when patient fells sick, falls or pushes a button.

The Standard and Interoperable Satellite Solution to Deploy Health Care Services Over Wide Area (HEALTHWARE) project (Loghelongue, 2007), is an integrated project of the Aeronautics and Space thematic priority of the 6th Framework Program (FP6) for satellite telecommunications systems and telemedicine applications. Thanks to the digital video broadcasting - return channel by satellite (DVB-RCS) technology, that offers satellite reception and transmission capabilities from anywhere, the project aims at developing and validating DVB-RCS based telemedicine solutions. It will focus on the areas of chronic respiratory disease, cardiology and oncology, through four main applications: medical training, tele-consultation, second opinions and monitoring and remote assistance at home.

Even if these systems are very interesting, none of them owns the characteristics presented in Section 1, primarily the possibility to be used by untrained people to perform single exams rather then continuous monitoring, with an immediate visual confirmation of the quality of the signal measurement. None of them could be used by the patient to control his/her health state through simple measurements even without sending the exam to a remote care center.

3 THE PROPOSED DVB-T TELE-HOME CARE SYSTEM

Digital Television (DTV) uses digital encoding techniques to broadcast video, audio and data contents to a receiver in the consumer's home (set-top box). In order to add multimedia information to the normal television program, the Multimedia Home Platform (MHP) allows the user to actively interact with the TV (Interactive TV) obtaining useful services and information. Based on the JavaTV platform, an extension and standardization of existing Java APIs in the context of DTV, MHP enables the interaction between the interactive applications and the set-top box by a software interface (middleware). Over a Virtual Machine, JavaTV applications (called Xlets) run in the set-top box with the support of a real-time Operating System (OS), which provides all the functionalities required including hardware resources access.

3.1 System Overview

The proposed system is depicted in Figure 1. The patient must own a TV, a DVB-T set-top box (with its remote control) connected in uplink to the telephone line, a personal smart card and a simple Base Station unit for bio-signals acquisition. The *Remote Care Center* (RCC) instead, has only a simple PC acting as TCP/IP server, since the patient's set-top box can send through the Internet the result of the exams.

The *Base Station* is a simple microcontroller-based acquisition unit to perform the digital biosignals acquisition. It is connected to the set-top box for data visualization and transmission, and it is controlled by it and then indirectly by the user *only* through the set-top box remote control, then simplifying the overall procedure.

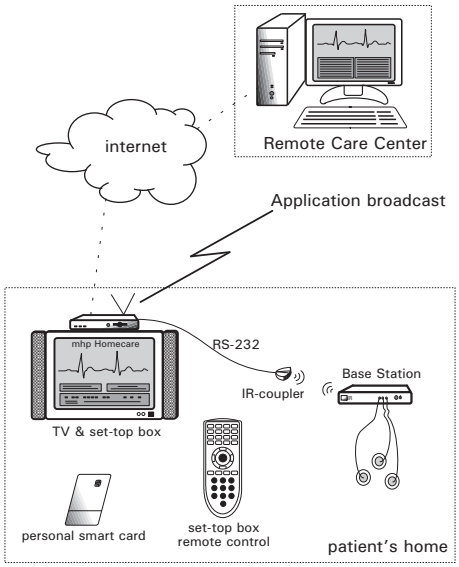


Figure 1: A schematic representation of the proposed DVB-T system for tele-home care.

The *smart card*, which must be programmed by the RCC, is used to identify the patient providing all the information needed to carry out the proper exams sending them to that RCC. The adoption of the smart-card for user authentication and to provide to the Xlet the required information about internet connection and RCC server IP address, avoids the patient to keep in mind or annotated elsewhere such information, improving the system usability. It also enables the personalization of the Xlet for the single user, since the exams that can be performed are only those programmed in the smart-card by the care staff. Since the largest part of DVB-T set-top boxes has a smart card reader (normally used for the pay-per-view services), no hardware on the base station is required to add this functionality.

The DVB-T MHP *set-top box* is like a computer, characterized by an input channel from user (the remote control), an output channel for the user (TV screen) and some ports to allow its interfacing with other devices, namely a serial port (RS-232) and an RJ-11 connector for the telephone line (since the set-top box has an internal modem), beyond the traditional audio/video connectors. The application presented in this paper has been tested on a Telesystem TS7.4DT set-top box, with the version 21p1 of the producer's software, and implementing the MHP 1.02 profile with some enhancements, primarily the addition of java packages for both the smart card and the serial RS-232 port management. The real-time OS is Osmosys. To interact with the base station, the set-top box uses the RS-232, whereas to send data to the RCC the integrated modem is used. In the proposed system the connection is in dial-up, so that it is surely simpler to have this facility in every home compared to broadband connections. The Xlet application is loaded into the set-top box by means of a broadcast transmission provided by a broadcaster, so that successive releases of it can be updated without the user's intervention.

3.2 Patient's User Interface

The graphic user interface (GUI) of the Xlet (hereafter called MHPHomecare) is user-friendly and intuitive, so that the patient can easily control the application by means of the remote control. The user can explore the application moving through the different full-screen frames (hereafter called FSFs), choosing among the possible options and functions by means of the remote control keys, as if it would be a PC keyboard. Three of the full-screen frames are generic whilst there is another FSF for each exam. The three generic FSFs are:

- the *primary* FSF, the first FSF shown when the application is started. It waits for the smart card insertion in the set-top box for user authentication and then it shows the patient's name (storing the other authentication data only for the transmission purposes), waiting for patient's commands. It is possible to close the application only from the primary FSF.
- the *about* FSF, that shows the application credits and can be launched only from the *primary* one;
- the *reduced* FSF, that corresponds to the window minimization to allow the patient to watch a TV program while the application is running in background. It can be launched from every FSF.

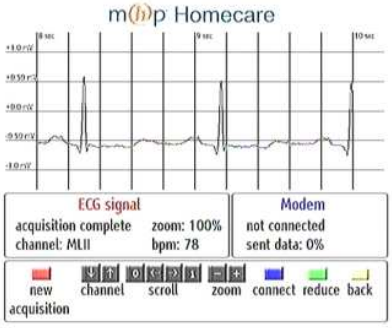


Figure 2: The ECG exam screen of the MHPHomecare application (ECG Exam FSF). The image has been acquired by means of a frame grabber from the set-top box video output, resulting in a poor quality and in altered colors.

In this application we used a simple 1-lead ECG to test the whole system, so there is only another FSF (a screenshot of this FSF is shown in Fig. 2), *ECG exam*. It enables the ECG acquisition and shows in real-time the samples acquired from the Base Station and all the other information sent by it about the current exam, such as the hearth rate (in bpm) and a warning message in case of poor quality signal. Once the exam has been saved, from the same FSF it is possible to send it to the server of the RCC, to review it on the TV (with zoom and scroll functions), to save a new exam or to return to the *primary* FSF. The patient can interact with the application moving through the different FSFs as depicted in Fig. 3.

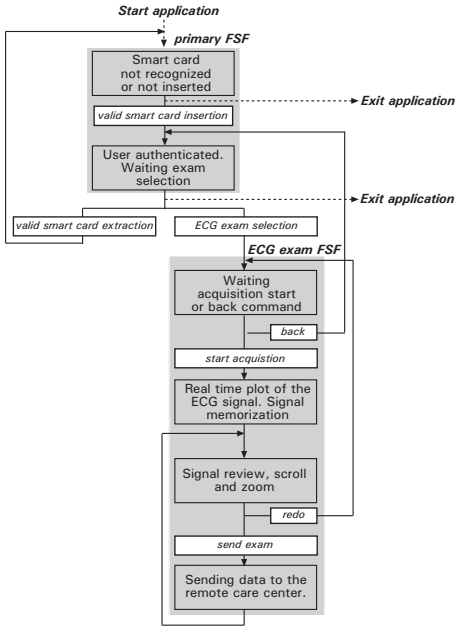


Figure 3: The interaction scheme between the patient and the set-top box application.

4 A PROTOTYPAL 1-LEAD ECG BASE STATION

In order to test our system we realized a simple microcontroller-based Base Station to acquire the patient's bio-signals. The Base Station is a black-box for the patient, and he cannot interact with it beyond the on/off power supply switch. The Base Station is battery powered, for simpler use and improved patient safety, and can include a different number of sensors and acquisition circuitry to serve different remote examination needs.

In the prototypal version presented in this work, it implements a single channel electrocardiogram (lead I) and consists in a classical ECG amplifier coupled with a very simple and low cost Digital Signal Controller (DSC), i.e. the Microchip^{TM} dsPIC30F4013. This is a 16-bit 30MIPS DSC enhanced with DSP hardware, such as a 17-bit x 17-bit multiplier, a 40-bit ALU, two 40-bit saturating accumulators and a 40-bit bidirectional barrel shifter. It also provides several embedded peripherals such as a 12-bit successive approximation analog to digital converter, and an USART, both of them employed in this prototype. The Base Station scheme is depicted in Fig. 4

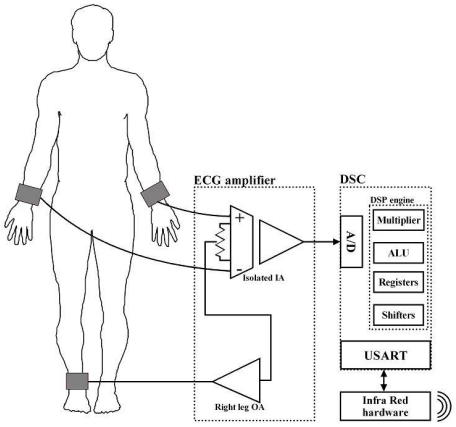


Figure 4: A schematic representation of the prototipal 1-lead ECG base station.

The fundamental characteristics of the Base Station (frequency bandwidth, sampling frequency and the number of bits of the A/D converter) are chosen according to the recommendations of American Heart Association for standardization and specifications in automated electrocardiography (Bailey et al., 1990). The ECG amplifier is a low-pass system with a variable gain up to 2000 and a cutoff frequency of 100 Hz. It consists of an optically isolated instrumentation amplifier with high CMRR connected to the wrists of the patient and a simple inverting operational amplifier connected to the right ankle for electromagnetic coupling noise rejection purpose, so that only 3 disposable (or reusable) electrodes are needed. The DSC is programmed to sample the ECG signal at 250 Hz, and it is interfaced through its internal USART to the DVB-T set-top box.

According to the international recommendations about the patient's safety, the physical connection between the Base Station and the set-top box is represented by an infrared (IR) link. The Base Station embeds a module for the IR transmission of the serial stream, whereas the set-top box exploits an external IR coupler connected to the RS-232 serial port.

In order to reduce both the residual 50/60 Hz en-

vironmental electrical noise we have implemented a digital notch filter, whereas a high-pass digital FIR filter (cutoff frequency 0.05 Hz) was added for the baseline drift elimination. All these computations are performed in real-time on the Base Station, since the set-top box is quite slow in processing due to the Java VM and to the applications running on it. On the Base Station has been also implemented a simple QRS detector for heart rate calculation and visualization on the TV screen (Friesen et al., 1990). The RR interval measurement is performed sample by sample, and if it is too far from the standards a warning code is sent to the set-top box to ask the user to control the correct connection of the electrodes. The average heart rate is sent when the acquisition stops and is referred to the average RR during the acquisition window.

5 THE INTERACTIVE APPLICATION: THE MHPHOMECARE XLET

MHPHomecare is a standalone application composed of several threads. It enables the non-simultaneous acquisition from the Base Station of an ECG exam, its visualization on the TV screen, and its sending via modem to the RCC where a server receives and stores the exam.

5.1 Graphic User Interface Thread

Taking into account the differences between PC and TV graphical applications, beyond the Java Abstract Window Toolkit (AWT), also the HAVi (Home Audio/Video Interoperability) API for graphic interfaces have been used (Java, 2007). We used Tiresias as default font, which has been created specifically for the TV screen, providing high readability and basing on the Western European ISO- 8859-1 fonts set.

A specific thread is in charge to create the objects related to the graphic elements composing the different FSFs of the application, and it shows/hides them depending on the current FSF.

5.2 Smart Card Management for User Identification

The smart card is read by means of the SATSA (Security and Trust Services API) classes (Satsa, 2007), that represent the new standard for smart card access in Java. To access the smart card, we used the methods based on the APDU (ISO7816-4) protocol, but the byte constituting the APDU strings are relative to the ACOS2 microprocessor card by Advanced Card Systems Ltd. To read the information stored into the memory of the card it is necessary to provide the CXS code to the method that opens the APDU connection, since the smart card is not a JavaCard (Dvt, 2006). A User Data File on the smart card stores the patient's name and its Personal Identification Code, the treating physician's name and its Identification Number, a permission exam code, some information about the ISP (user ID, password, telephone number) and the server IP address.

A thread, that is active in the primary FSF, verifies the correct insertion of the smart card into the reader, hence reading the User Data File. If the card is not properly inserted and recognized, the Xlet remains blocked on the primary FSF. Once the card has been recognized, the permission exam code is used to allow the user to perform only some specific exams among those available, hence disabling all the other ones.

5.3 Serial Port Management for Exam Acquisition

The serial port control is possible by means of the `it.dtt.comm` package, available on the set-top box used in this application. The communication follows analogue principles we can find in `javax.comm`, and the serial port behaves according with the RS-232 standard. A thread manages the communication through it.

The set-top box sends to the Base Station the commands to start the acquisition and to stop it. Both such commands are simply 1-byte words. The "start" command word is logically divided in two parts, i.e. the actual command to enable the serial port transmission, and a code that identifies the exam the user chose by means of the Xlet application. The set-top box receives from the serial port the samples of the current exam in frames composed of N samples each, where N changes from exam to exam as a function of the sample rate. The application deals with 8 or 16-bit signals (the data are stored in arrays of short). In the current implementation only the ECG is available on the Base Station, and it is acquired at 250 sample/sec, 16 bit/sample. With such sample rate and data width, the maximum re-painting rate of the graphic area where the signal is plotted is 200ms, which in turn means that a frame of input samples must consist of minimum N = 50 samples.

After the exam stops with a "stop" command word, the Base Station sends to the set-top box a last word with the value of the heart rate expressed in bpm so that the Xlet updates with this information the

proper FSF.

5.4 Modem Management for Exam Transmission

A specific thread is in charge to manage the modem, and it operates by means of the `org.dvb.net.rc` package (Java, 2007), jointly with the `java.net` and `java.io` packages. The operations performed by means of these packages are: ISP number dialing, ISP authentication and TCP/IP data transfer towards the server by means of a socket opening.

The user can start the transmission to the RCC from the exam FSF simply pressing a key on its remote control. In response to this action, the thread establishes the connection with the ISP using the information previously read from the smart card. Once the connection with the ISP has been established, the thread opens a client socket specifying the IP address of the server of the RCC. This IP address is read from the smart card too, since different users need to send their exams to different RCCs. At this point, the thread sends a stream composed of two main parts: the patient's data and the exam data (collected during the acquisition phase). Once the whole data stream has been sent, the connection is closed and another exam can be acquired.

6 CONCLUSIONS

In this paper, a DVB-T framework for tele-home care was presented along with its prototype implementation.

Thanks to the DVB-T set-top box, which is very similar to a computer with the possibility to be programmed, the system can interact with external custom peripherals and can be connected to the web. Easy to use by people not skilled in digital electronics, and since tele-home care is mainly used to monitor elderly people which eventually falls in the previous category, pushed by the imminent deadline de fined by the UE for the switch-off of the old analog broadcasting transmission system, such tele-home care framework could represent the best solution in terms of quality and costs. Compared to traditional systems, the presence of a visual environment on the TV screen allows a more friendly use providing also more detailed information and feedbacks about the signal quality, and guiding the user through all the exam procedure without any required printed manual.

Experimental results show the high potentiality of the proposed solution and deserve further improvements such as the development of a more powerful base station (or a set of different base station models), the realization of a digital broadcasting system, and the extension to the official sanitary card, currently under trial in some Italian regions and in other European countries. It is actually under test a complete Visual application to read and analyze the exams on the Remote Care Center.

ACKNOWLEDGEMENTS

This work is part of the *Personal e-Care* project, which has been partially founded by INFN, the Nuclear Physics Italian Institute.

REFERENCES

Bailey, J., Berson, A., Garson, A., Horan, L., Macfarlane, P., Mortara, D., and Zywietz, C. (1990). Recommendations for standardization and specifications in automated electrocardiography: bandwidth and digital signal processing. In *Circulation*. American Heart Association.

Dvt (2006). Compatible dtv receivers for the italian market: baseline requirements - DGTVi D-Book 1.1 (rev.4). DGTVi.

Friesen, G. M., Jannett, T. C., Jadallah, M. A., Yates, S. L., Quint, S. R., and Nagle, H. T. (1990). A comparison of the noise sensitivity of nine QRS detection algorithms. *IEEE Transactions on Biomedical Engineering*, 37(1):85–98.

Java (2007). Javadoc for MHP specification. http://www.mhp.org/mhp_technology/mhp_1_0.

Kyriacou, E., Pavlopoulos, S., Koutsouris, D., Andreou, A., Pattichis, C., and Schizas, C. (2001). Multipurpose health care telemedicine system. In *Proc. of the 23rd Annual EMBS International Conference, Istanbul, Turkey*.

Loghelongue, P. (2007). Satellite based telemedicine: Accessibility, quality and safety. In *Proc. Tromso Telemedicine and eHealth Conference, Tromso, Norway, June*.

Mailhes, C., Castanié, F., Henrion, S., Lareng, L., Alonso, A., Weber, J.-L., Zeevi, B., Lochelongue, P., Depeursinge, Y., Kollias, V., and Ferhaoui, M. (2003). The U-R-Safe telemedicine project: improving health care of the elderly. In *Proc. of MIE2003, St.Malo, France*.

Pierucci, L. and DelRe, E. (2000). An interactive multimedia satellite telemedicine service. *IEEE MultiMedia*, 07(2):76–83.

Satsa (2007). Security and trust services API for J2ME (SATSA); JSR 177. http://java.sun.com/products/satsa.

ECGAWARE: AN ECG MARKUP LANGUAGE FOR AMBULATORY TELEMONITORING AND DECISION MAKING SUPPORT

Bernardo Gonçalves, José G. Pereira Filho
Computer Science Department, Federal University of Espírito Santo (UFES), Vitória, Brazil
bgoncalves@inf.ufes.br, zegonc@inf.ufes.br

Rodrigo V. Andreão
Electrical Engineering Department, Federal University of Espírito Santo (UFES), Vitória, Brazil
rodrigo@ele.ufes.br

Keywords: Telecardiology, Telemonitoring, ECG data, Interoperability, Context-Awareness, Emergency and decision support.

Abstract: The ambulatory electrocardiogram (AECG) can be acquired and transmitted through mobile and wireless technologies and devices to foster heart's telemonitoring anytime, anywhere. This sort of service is purposeful when combined with ECG analysis systems and infrastructural support for providing context-aware services. Such setting makes efficient emergency services possible as well as improves the support to physician's decision making. This paper presents an ECG XML-based markup language that extends ECG reference standards in order to cover patient's heart telemonitoring during his/her daily activities. The ECG data format we propose is then applied in a real scenario.

1 INTRODUCTION

The rapid expansion of ICT has been allowing the creation of new services on Healthcare. The Telecardiology, in particular, has developed itself mainly through the transmission of the electrocardiogram (ECG). On one side, the ECG is fast, cheap and non-invasive when compared with other cardiology examination procedures. On the other side, the analysis of the ECG waveform can identify a wide range of heart illnesses, which are distinguished by specific modifications on ECG elementary waveforms. These are the reasons why ECG is the most frequently applied test for measuring heart activity in Cardiology. According to estimates, more than 100 million ECG's are recorded yearly in Western Europe (Fischer and Zywietz, 2003).

The storage and transmission of all these data have then been object of some initiatives concerning ECG format standardization. The oldest standards are AHA/MIT-BIH (Goldberger et al., 2000) and SCP-ECG (SCP, 2002), regarding records' storage and transmission respectively. In face of the Internet popularization, however, novel standards have been conceived in order to integrate interoperable and user-driven solutions, standing out FDADF (Brown et al., 2002) and ecgML (Wang et al., 2003), both based on the XML markup language.

Nonetheless, such standards do not take into account heart's telemonitoring, which calls for the representation and transmission of the ambulatory electrocardiogram (AECG). The portable device that records the AECG was invented by Norman Holter in 1957. Since then, the ICT advances in addition to improvements in the accuracy of ECG software-based analysis systems have opened new potential uses to the AECG. Indeed, it is largely employed by the medical community, mostly for diagnosis and/or therapeutic treatment of the myocardial ischemia, which constitutes a pre-infarct. Since most ischemia episodes are related to increases in heart rate possibly associated to day-to-day variability of physical or emotional activities, AECG is indicated for patient's heart monitoring throughout his or her daily activities. With this in mind, the most suitable duration of a recording session to detect and quantify ischemia episodes is probably 48 hours. Some experiments point out that most patients are quite comfortable wearing the recorder for 48 hours (Crawford et al., 1999).

Besides, with the advent of a new Computing paradigm, the Pervasive Computing, context-aware

systems provide new features, standing out for collaboration among professionals, systems, and action triggering from the detection of changes in the context of the user. This can be verified with the growth of initiatives dedicated to patient's monitoring, whether at home or in emergency situations wherever they take place. An example is the Awareness project (Awareness, 2007). Such efforts have taken advantage of the latest advances in mobile and wireless technologies and devices, in general, and may rely on signal processing algorithms, in particular, in order to provide both alarms generation and decision making support.

In face of all these aspects, we advocate that an ECG data format should cover the particular issues concerning telemonitoring through AECG, such as the ones related to emergency assistance (e.g. patient's location), or the activities performed by the patient during the ECG recording session (e.g. rest, physical exercise, routine activities, etc.). In this article we propose a novel ECG data format that extends former reference standards in order to cope with real-time telemonitoring and decision making.

The paper is organized as follows. Section 2 discusses the background of ECG data standards as well as aspects of telemonitoring; Section 3 presents the ECG data format we propose in this paper; Section 4 introduces a usage scenario where the format proposed is applied; and, finally, Section 5 concludes the paper and depicts future work.

2 BACKGROUND

Throughout the last thirty years, we can notice a regular evolution of standards regarding ECG record's representation and transmission. One may state that each standard resembles its purpose and the ICT environment at the time of its arising.

Since 1975, the Massachusetts Institute of Technology (MIT) together with laboratories of the Beth Israel Deaconess Medical Center have carried out research concerning medical examinations analysis and related points. As a result, in 1980 the MIT-BIH Arrhythmia Database was deployed after testing and standardization for arrhythmia detection and evaluation. Also at this time, the American Heart Association (AHA) has deployed the AHA Database for Evaluation of Ventricular Arrhythmia Detectors (Goldberger et al., 2000). Together, those databases have been largely used and played an important role on research in the field of Cardiology (Moody and Mark, 2001).

The AHA/MIT-BIH standard has focused on bringing in an ECG records' library for providing input for developers of ECG analysis systems. It, in fact, is responsible for substantial advances on ECG data processing. This standard, however, does not aid interoperation over the Internet due to its tight coupling with programming language. Moreover, it is not human readable, which is a desirable requirement with respect to the analysis of electrocardiography's domain experts.

Later, there was a great effort to conceive the SCP-ECG - Standard Communications Protocol for Computer-Assisted Electrocardiography (SCP, 2002). SCP-ECG is a specification concerning ECG data format as well as transmission procedure from the acquisition device to the host where the message is stored and retrieved. From 1989 to 1990, it was carried out a survey on ECG compression methods that has as a result led to an original approach for signal compression (Fischer and Zywietz, 2003).

Nonetheless, despite the SCP-ECG allows suitable data compression, the elements of the format are defined at the bit level. This obstructs changes on the format, either for updating or customization, as well pushes final applications (i.e., ECG viewers) to be familiar with SCP codes. As per (Clunie, 2004), SCP implementation is an awkwardly task especially on the compression mode. Considering that computational resources are currently more accessible than at the creation of the SCP, bandwidth over the networks, memory capacity and disc space are not main concerns as they were before. Meanwhile, other concerns have taken place on Telemedicine scenarios, such as the need for platform- and application-independent solutions involving human readable data models.

In this trend, and also as a result of the grown popularity of the Internet, XML-based formats as FDADF and ecgML have been increasingly used on Telecardiology research. The Extensible Markup Language (XML) has played an important role on data exchange over the web, especially by providing the separation of data content and presentation. After XML has became a W3C recommendation in 1998, several domain specific languages were created from a XML Schema. In this way, several committees of Health organizations such as CEN/TC251, Health Level Seven (HL7), American Society for Testing and Materials (ASTM), etc, have worked on the development of recommendations for using XML on Telemedicine research.

The Food and Drug Administration (FDA) has carried out a survey on ECG standards and has chosen the XML technology for data representation based on the HL7 ECG annotation message v3 (HL7, 2003). As a result, in 2002 it has produced the FDA XML Data Format (FDADF). The FDADF is

an effort to reach the standardization of ECG data representation for all stakeholders share the same view (Brown et al., 2002). Looking for addressing requirements previously defined, the scope of the FDADF specification covers ECG data as much as significant submission information. FDADF has achieved a significant progress on ECG data representation by using XML. Nevertheless, as per (Wang et al., 2003), it does not exploit as far as possible XML features. That is because, on account of ECG viewer applications' concerns, it has incorporated elements related to data presentation in its metamodel, rather than to cope only with data content.

More recently, in 2003, the ecgML was developed in face of the increased demand for a standardized application- and platform-independent ECG format. This one has been conceived from the former standards (especially the FDADF), reusing then concepts and nomenclature. The ecgML allows ECG data analysis and transmission between heterogeneous platforms (Wang et al., 2003). Indeed, rather than FDADF, the ecgML has comprised only data content. It holds benefits such as flexibility, readability and descriptiveness. Nonetheless, as remarked by the authors themselves, there are issues left to evaluation, such as concepts still not covered in ecgML (Wang et al., 2003).

As a matter of fact, the more is the emergence of new technologies increasing the usage potential of computer systems, the more there are usage scenarios foreseen. As a result, further information can be explored promoting then more useful services. In this way, a data format for wrapping biomedical signals, in fact, constitute an interface between data acquisition and data usage systems (see Figure 1). Therefore, such a data format should not be a restrictive mean for useful data acquired from sophisticated devices. As opposed, it should abstract the complexity related to biomedical signals acquisition to the health professionals' environment. This concern is particularly worth in context-aware telemonitoring of patients' heart relying on both wireless and mobile technologies and devices and the transmission of AECG. The existing ECG standards, however, lack this concern and neither were conceived from advanced modeling techniques such as domain ontologies.

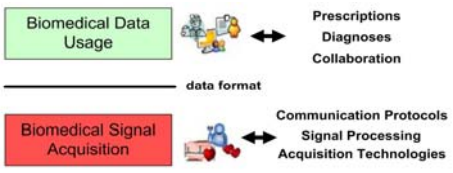


Figure 1: Separation between data acquisition and usage.

As an effort to cover this gap in literature, we have carried out an extensive research on the Electrocardiography domain. At this time, we have developed an ECG domain ontology which is presented elsewhere (Gonçalves et al., 2007), and the ***ecgAware***, a XML-based ECG data format which is the focus of this article. The *ecgAware* extends former standards especially by covering AECG aspects related to context-aware telemonitoring. On the next section we elaborate on the *ecgAware* markup language, remarking the main issues we have previously mentioned.

3 ECGAWARE

The *ecgAware* has a tree hierarchical structure which is described in the following in a prefix way, i.e., expanding each significant XML complex element on the left. The main elements are depicted on diagrams in the figures 2 to 5. XML elements and attributes are both referenced in bold and italic, (the elements have the first letter capitalized); optional elements or attributes are dotted in the diagrams.

The *ecgAware* model constitutes an ***ECGStudy*** (see Figure 2) of a single patient, which integrates attributes that provide some prior data. These data are ***studyID***, a unique ID; ***studyTimeStamp***, i.e., date and start time of the latest ECG record present in the message; ***dateTimeZone***, which supplies the acquisition local time zone (based on SCP); ***studyLocation***, holding the latest location obtained; the ***alarm*** attribute indicates that at least one record inside the study contains an abnormal event, which may be either detected by an ECG analysis system or triggered by the patient. In case it is flagged true, the *ecgAware* message supports an efficient emergency service by the ***studyTimeStamp*** and ***studyLocation*** attributes. Finally, ***computerID*** identifies the machine where signal processing takes place (based on SCP) and ***investigatorID*** is a unique ID of the health professional which blames for the ECG study (based on FDADF). ***ECGStudy*** has three child elements: *(i)* ***PatientData***, for patient's demographics data and electronic record; *(ii)* ***Record***, the ECG record produced in each recording session; and *(iii)* ***Comments***, for free text.

The ***Demographics*** element then comprises data for identifying and contacting the patient (inspired on ecgML); its child elements are ***Name***, ***Sex***, ***DOB*** (date of birthday), ***Address***, ***Phone***, ***Fax*** and ***Email***. Meanwhile, ***EPR*** represents a basic patient's electronic record; it is composed by patient's ***Height*** and ***Weight***; the boolean elements ***Hypertension***, ***Diabetes***, ***Smoker*** and ***Alcohol***; ***Other*** for inserting

other clinical data; and ***Comments***, a free text field. ***Demographics*** and ***EPR*** may be obtained by means of a simple anamnesis. Those data are optional because there may be situations (e.g. an emergency) where there is no time for collecting them.

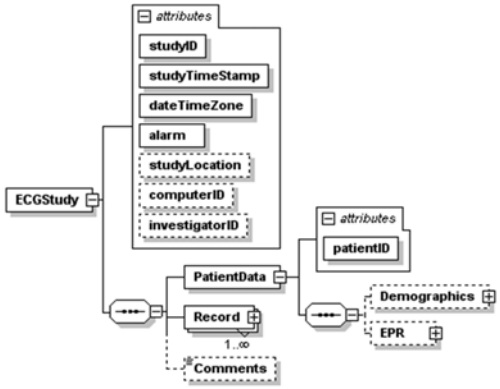


Figure 2: The ***ECGStudy*** root element.

ECG data are laid in the ***Record*** element (see Figure 3). A ***recordID*** attribute identifies the record; The ***RecordingDevice*** element describes the acquisition device used to obtain the record (based on FDADF) and filtering technique(s) performed by it (based on SCP); ***RecordingSession*** bears the recording session context, and is especially useful for emergency services and decision support; ***RecordChannel*** (min. one, max. twelve) constitutes the ECG signal acquired through a channel; ***GlobalAnnotations*** and ***GlobalMeasurement*** in turn (inspired on FDADF and SCP respectively) are annotations and measurements related to all leads; and lastly, ***Report*** is a record finding carried out either by a physician that interacts with a system or by an analysis system to be further verified by a confirming physician.

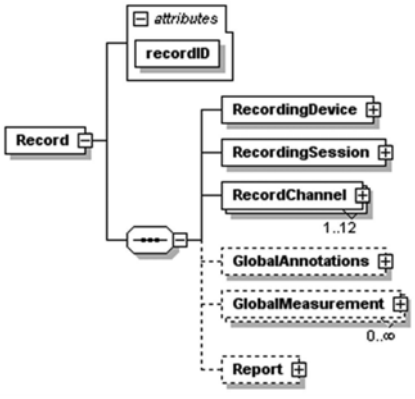


Figure 3: The ***Record*** element.

RecordingDevice has a ***deviceID*** to identify the device by a serial number. It has also a ***Type***, a ***Manufacturer*** and a ***Model*** (e.g. Holter, Space Labs, 90205). The ***BaselineFilter***, ***LowpassFilter*** and zero or many ***OtherFilter*** elements constitute noise filtering to overlook signal frequency components over superior bounds and other filtering possibly performed on the signal, respectively.

As we previously mentioned, patient's heart continuous telemonitoring can support diagnosis and/or therapeutic treatment of the myocardial ischemia. This is possible by means of a long-term AECG ***RecordingSession*** (see Figure 4). With this in mind, we included the ***Activity*** element, which lays up a description of each activity performed by the patient during the recording session (e.g. rest, physical effort, etc). This information can be either obtained by user interaction with the ECG acquisition system (in replacement of the paper in which patients used to populate his/her activity/time over the recording session); or much better, acquired by a sensor device such as a video camera jointly with an eye-tracking system (Zhai, 2003), or by other sensing techniques (Boudy et al 2006).

Still from a context-awareness standpoint, the patient's context during a recording session may be used, for example, to guide an ambulance to the patient's location whenever an emergency takes place. This sort of feature is possible thereby small mobile devices which permit, nowadays, patient's vital signs and location telemonitoring even in outdoor scenarios. That is why we included the ***AcquisitionLocation*** element, which holds the latest patient's location acquired in a ***RecordingSession*** from a device such as GPS.

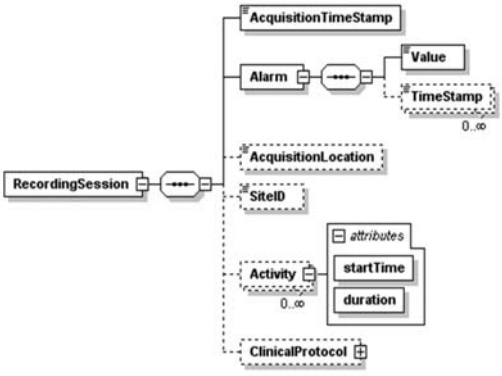


Figure 4: The ***RecordingSession*** element.

Besides, ***AcquisitionTimeStamp*** hands over date and start time of each session; ***Alarm*** flags on a true or false ***Value*** for abnormal event(s) either identified by an ECG real-time analysis system or triggered by the patient, and keeps on zero or many ***TimeStamp*** elements date and time of the detected event(s). In case ***Alarm*** is true, on one side, an *ecgAware*

message comprising a partial record must be transmitted as far as the abnormal event was detected; on the other side, the whole ECG must be recorded including all alarm events occurred during the recording session related to it. Lastly, ***SiteID*** is an abstract description of the place whereby the session took place (e.g. domicile).

ClinicalProtocol, rather than in other ECG data formats, is placed in the ***RecordingSession*** element. Indeed, it is related to the meantime of a session instead of a range of sessions. It then is composed by ***DiastolicBP*** and ***SystolicBP***, taken in the session at some ***timestamp*** under a value ***unit***; ***Medication***, specifying drugs which the patient has been using; and finally, ***Sweaty*** and ***Pale*** (based on ecgML) indicating true or false for abnormal sweat and abnormal looking skin on the face, respectively.

The ECG signal is obtained from correlated observation series taken at the same time by electrodes placed on some positions on the human body. These placements, when combined, provide different viewpoints of the heart electrical activity, i.e., the ECG leads. In Electrocardiography twelve leads were standardized. ECG data is thus laid up on one or more (max. twelve) ***RecordChannel*** elements (see Figure 5) standing for the leads. The ***Channel*** element identifies the lead (e.g. Lead II); ***Waveform*** contains the XY signal; ***ChannelAnnotations*** and zero or many ***Measurement*** elements, in turn, are annotations and measurements related to a lead (all inspired on FDADF and ecgML).

The ECG samples are obtained from the observations performed by the device over the time, and thus constitute XY values. They are situated in the ***Waveform*** element by ***XValues*** and ***YValues***. However, since observations are evenly spaced in time, we do not need to store time values (***XValues***) in the XML document. They rather can be easily obtained by the ***Xoffset***, ***Duration*** (of the record) and ***SampleRate*** elements (all of them holding a ***unit*** attribute). The sample values (***YValues***), otherwise, must be covered in the XML Document, even though there are different options to get it done. Those values have also a ***unit*** and may be laid up either *(i)* in an external file, which the link path is indicated by ***FileLink***; or *(ii)* by an integer series ***IntValue*** (which can be easily converted to float by using a scale); or even *(iii)* by a binary encoding (***BinaryData***). Both ***IntValue*** and ***BinaryData*** are composed by the ***From***, ***To***, ***Data*** and ***Scale*** elements. They are respectively the beginning and ending of the waveform in the X axis, the sample values, and a scale factor to obtain the real number of each value. The ***BinaryData*** has also a data ***encoding*** attribute (e.g. Base64).

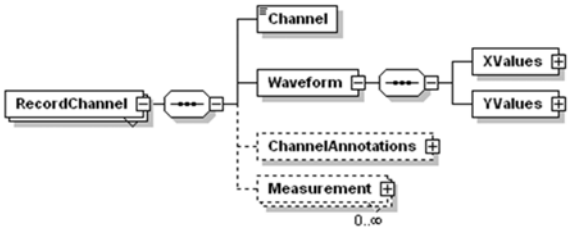


Figure 5: The ***RecordChannel*** element.

ChannelAnnotations mark significant events identified on the waveform. They are carried out by an ***author***, which may be either a system or a physician. Annotations may be either about points (***PointNotation***) or time intervals (***WaveNotaion***). The former involves a ***PointLabel*** describing the point, an ***XValue***, an ***YValue***, and a ***Comment***. The latter in turn marks beginning, peak and ending time values of one or more waves by the ***Onset***, ***Peak*** and ***Offset*** elements, respectively. Moreover, it holds also an ***Interpretation*** (e.g. abnormal) of the waveform in this time interval. It is worth to say the ***WaveNotation*** element is basically addressed by the elementary forms, or waves, which compose the heart beat. They were defined by Einthoven in 1895 as PQRST; we can abstract the elementary forms by ***Pwave***, ***QRScomplex*** and ***Twave***. Zero or many ***OtherWave*** elements may also be considered. On different leads, a specific elementary form can be viewed in a better or worse way, exception by the QRS complex, which can be well viewed through whichever lead. This is the reason why we made only the ***QRScomplex*** element required on ***WaveNotation***.

Besides annotations, zero or many ***Measurement***(s) can be made either of the duration or of the amplitude of elementary forms. We can distinguish that by means of the ***label*** and ***unit*** attributes, e.g. P-duration and ms (based on ecgML and SCP).

Global annotations and measurements, as opposed to the channel ones, are related to all leads. This sort of annotations then can be performed either by a physician marking a vertical line correlating the same ***XValue*** on all leads through an ECG viewer application, or simply by a system from an average of the correlated channel annotations. The ***GlobalAnnotations*** element (inspired on FDADF) thus discriminates itself from ***ChannelAnnotations*** only by, on the former, all elementary forms are required. The ***GlobalMeasurement*** element (inspired on SCP), otherwise, has exactly the same structure of the channel ***Measurement***. The former, however, is obtained from an average of the correlated channel measurements.

Finally, we have admitted a ***Report*** element to provide a finding about the ECG record. This report is carried out by an ***author*** which may be either a system or a physician that has saw the record through an application and then has edited it. The finding comprises ***HeartRate***, ***ElectricalAxis*** of the heart, and ***Diagnosis***.

4 USAGE SCENARIO

As part of a research program in healthcare and bioengineering technologies at UFES, in Brazil, we have been developed in the TeleCardio project a context-aware system for remote monitoring patients with cardiological syndromes (Andreão et al., 2006a). In TeleCardio, the patient can have his/her heart activity monitored anytime either in domicile, ambulance, or outdoor scenarios. The TeleCardio system carries out the transmission of the AECG in combination with contextual data (e.g. location) in order to allow physicians follow their patients' condition in real-time and report diseases remotely. TeleCardio, in fact, is a rich field for applying *ecgAware* to wrap and deliver ECG signals, related data and contextual data. The Figure 6 depicts the TeleCardio architecture.

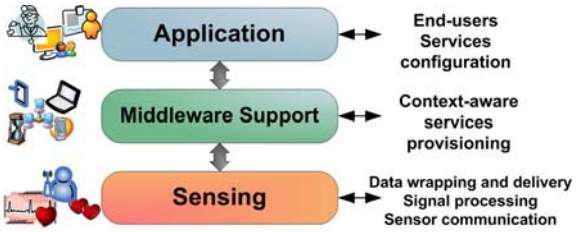


Figure 6: TeleCardio Architecture.

The Sensing layer comprises the *ECG Wrapper*, an integrated system (hardware and software) for acquiring the ECG signal from a Holter device. This system handles wireless communication with the device, signal processing, data wrapping and delivery. One of the components of such system is an ECG analysis software (Andreão et al., 2006b), which makes use of an enhanced approach for ECG classification and segmentation. This software thus produces the ECG enhanced data for populating the *ecgAware* model. The ECG Wrapper, in fact, hides all complexity related to biomedical signals acquisition for the middleware and application layers by delivering *ecgAware* data to them.

The middleware layer, named Infraware (Filho et al., 2006), provides context-aware services for supporting client applications. Examples of such services are: *(i)* to supply subscription management for the health client applications as well as to manage the interactions between these applications and the ECG Wrapper; *(ii)* to guarantee privacy and access control to patients, physicians, etc; *(iii)* to guide an ambulance from the patient's domicile to a hospital by choosing the best traffic routes; and *(iv)* interpretation of pieces of information in order to trigger emergency services, e.g. when ***alarm*** is flagged true.

The Application layer in turn addresses services configuration, users' profile and so on. The Figure 7 shows a web application that we have developed in the TeleCardio project whereby physicians can view the patients' ECG signals and take advantage of the *ecgAware* features. By means of such an application, the physician can follow his/her patient's heart activity anywhere, anytime. In case an emergency takes place, the physician is notified both by the application (in case he/she is online on the system) and by a SMS message his/her cellular phone.

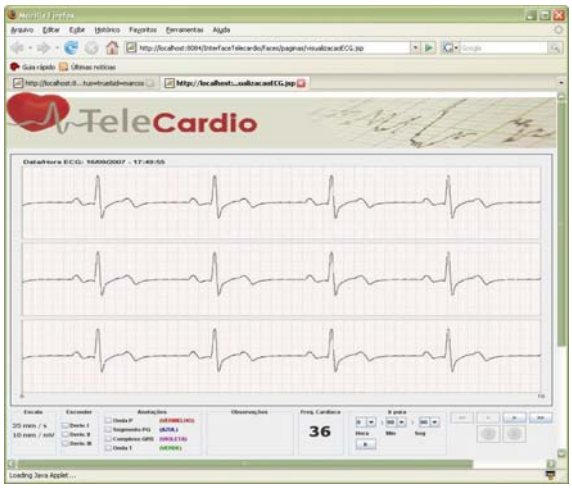


Figure 7: Snapshot of the TeleCardio's ECG viewer.

Because we are speaking of telemonitoring in real-time, the ECG Wrapper has then to pack the AECG record into small pieces of *ecgAware* data for delivery by each 30 seconds. This time interval is related to hardware optimal operation as well as the emergency procedure. In fact, one could argue that the XML format do not meet computational efficiency as much as a binary format. Nonetheless, we can overcome this either by using a binary file to store and transmit the ECG XY data (see the ***FileLink*** element) or by developing a compression procedure in order to reduce the XML file. In (Erfianto, 2004), for example, the compression scheme has reached a reduction of up to 53% for ECG data and 87,5% for patient data. The technique

used parses the XML document to an ASN.1 format and, in the sequel, to a binary-encoded format.

5 FINAL CONSIDERATIONS

Currently we have advanced wireless and mobile technologies and devices as well as systems that carry out biomedical signal's analysis through signal processing algorithms. In this work we have argued on the worth of taking advantage of such resources in order to improve emergency services and decision making support in the Healthcare domain. In this way we can make the acquired sensor data much more useful.

In the scope of Telecardiology, in particular, we have elaborated in this paper how the existing ECG data format standards lack this concern. We have then proposed a novel ECG model striving not only for application- and platform-independence and focusing data content, but also admitting elements to address telemonitoring concerns. As a result, we can provide better emergency services and decision making support, meeting then the requirements related to pervasive scenarios in Healthcare.

The first usage results remarks *ecgAware* is suitable for its purpose. After all, throughout the TeleCardio evaluation we will test such data format under an intensive usage by one or more medical communities. Hence, we will have statistical metrics to better evaluate it.

Former usage scenarios in Telemedicine, e.g. remote reporting, are still covered by the *ecgAware* data format. In fact, it embraces features of the former ECG reference standards. Future usage scenarios may cover other vital signs telemonitoring. This is feasible by using the same research methodology, i.e. exploring each biomedical signal domain as we did in this work with the ECG. Moreover, the XML technology provides flexibility such that we can incorporate several XML schemas to a root one. We thus can keep the elements regarding telemonitoring, in general, and to design a record structure for each biomedical signal we choose to bear.

ACKNOWLEDGEMENTS

This work has received financial support from FAPES (grant no. 31024866/2005) and CNPq (grant no. 50.6284/04-2). The first author would like to express his gratitude to Giancarlo Guizzardi for his invaluable contribution to the progress of this work.

REFERENCES

Andreão, R. V., Filho, J. G. P., et al., 2006a. TeleCardio – Telecardiologia a serviço de pacientes hospitalizados em domicílio. In *X Brazilian Conference in Health Informatics (CBIS'06)*, Florianópolis, Brazil.

Andreão, R. V., et al., 2006b. ECG signal analysis through hidden Markov models. In *IEEE Transactions on Biomedical Engineering*, v. 53, n. 8.

Awareness project, 2007. http://www.freeband.nl/.

Boudy, J., Andreão, R. V., et al., 2006. Telemedicine for elderly patient at home: the TelePat project. In *International Conference on Smart Homes and Health Telematics*. Belfast, Northern Ireland.

Brown, B., Kohls, M., Stockbridge, N., 2002. FDA XML data format design specification. Retrieved August 9, 2006 from http://www.cdisc.org/discussions/EGC/FDA_XML_Data_Format_Design_Specification_DRAFT_B.pdf.

Clunie, A., 2004. Extension of an open source DICOM toolkit to support SCP-ECG waveforms. In *2nd OpenECG Workshop*. Berlin, Germany.

Crawford, M., et al, 1999. ACC/AHA Guidelines for ambulatory electrocardiography. In *Journal of the American College of Cardiology*.

Erfianto, B., 2004. *Design of a Vital Sign Protocol Format Using XML and ASN.1*. M. Sc. Thesis, University of Twente, The Netherlands.

Filho, J.G. P., et al., 2006. Infraware: um *middleware* de suporte a aplicações móveis sensíveis ao contexto. In *Proceedings of the 24th Brazilian Symposium of Computer Networks*, Curitiba, Brazil.

Fischer, R., Zywietz, C., 2003. SCP How to Implement - Part I. Retrieved August 8, 2006, from http://www.openecg.net/.

Goldberger A., et. al., 2000. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. In *Circulation* 101(23):e215-e220.

Gonçalves, B., et al., 2007. An electrocardiogram (ECG) domain ontology. In *Proc. of the 2nd Workshop on Ontologies and Metamodels for Software and Data Engineering (WOMSDE'07)*, João Pessoa, Brazil.

HL7 ECG Annotation Message v3, 2003. Retrieved August 9, 2006, from http://wwwhl7.org/V3AnnECG/index.htm.

Moody, G., and Mark, R., 2001. The impact of the MIT-BIH arrhythmia database. In *IEEE Engineering in Medicine and Biology Magazine*.

Standard communication protocol – Computer-assisted electrocardiography N02-15, 2002. CEN/TC251 Secretariat. Retrieved August 8, 2006, from http://www.centc251.org/.

Wang, H., Azuaje, F., Jung, B., Black, N., 2003. A markup language for electrocardiogram data acquisition and analysis (ecgML). In *BMC Medical Informatics and Decision Making* 3:4.

Zhai, S., 2003. What's in the eyes for attentive input. In *Communications of the ACM*, 46(3).

INTRODUCING A MOBILE SYSTEM FOR THE EARLY DETECTION OF CARDIAC DISORDERS AS A PRECAUTION FROM A CARDIOLOGISTS' VIEW

Evaluation of a Survey

Onur von Burg, Marco Savini, Henrik Stormer and Andreas Meier
University of Friborg, Boulevard de Prolles 90, 1700 Fribourg, Switzerland
[onur.vonburg, marco.savini, henrik.stormer, andreas.meier]@unifr.ch

Keywords: Cardiology m-health e-health mobile survey heart.

Abstract: This paper illustrates the results of a survey of practicing cardiologists that were asked about various aspects of a simplified telecardiology scenario using mobile devices. Such devices are becoming ubiquitous assets in everybody's life. Their application in a healthcare environment aims not only at supporting the patients over traditional consultations but also maximizes the content of their health status information. The results of the survey may help application developers to focus their efforts of applications in a similar setting.

1 EHEALTH AND MHEALTH

(Haas, 2006) defines eHealth as

> ...the application of Internet and other related technologies in the health care industry to improve access, efficiency, effectiveness, and quality of clinical and business processes used by health care organizations, practitioners, patients, and consumers to improve the health status of patients.

mHealth extends this notion by considering mobile actors (patients, health professionals, ...) and their environment. (Istepanian et al., 2006) simply define it as "emerging mobile communication and network technology for health care". Certain areas of eHealth like monitoring over distance can be greatly improved when considering ubiquitous mobile devices. (del Pozo et al., 2006) defines that potential needs should meet certain criteria so that:

- eHealth services should be provided at any time and any place, depicting the need for efficient, ubiquitous and secure institutional care.
- Efficient coordination tools should be provided to all health professionals dealing with each single patient, in order to allow the implementation of patient centered care.

The recent advances in mobile technology like higher bandwidths or powerful smartphones with advanced features allow the creation of new medical applications that support both patients and doctors.

A number of researchers have worked on the idea of assigning mobile devices to patients. It is possible to distinguish between the following three domains:

1. Mobile devices are used to help the patient by providing information.
2. Mobile devices are used to transmit physiological parameters.
3. Mobile devices are used to alert patients or medical professionals when certain physiological parameters become critical.

The dialogue-based monitoring system in (Komninos and Stamou, 2006) aims at supporting elderly people in their preferred environment and can therefore be assigned to the first domain.

The MOEBIUS project (Mobile extranet-based integrated user services) that integrates doctors and patients by submitting different physiological parameters (Fischer et al., 2006) can be considered as an application of the second domain. Similarly, (Leimeister et al., 2005) describes the usage of mobile devices in order to assist young cancer patients and concludes that the usage of such a system has a number of advantages: higher compliance of appointments with alerting functionality, higher data quality, less work for the doctor to prepare the documentation as well as less errors in the documentation. The use of context

in a mobile healthcare scenario using the information groups Who, Where, When, Why and What has been described in (Savini et al., 2007).

An example of the third domain collecting real-time electrocardiogram signals including basic arrhythmia detection with automatic alerting to a call center is illustrated in (Liszka et al., 2004). Their system architecture uses readily available commercial off-the-shelf components.

1.1 Cardiology

1.1.1 Overview

Cardiovascular reasons represent the largest cause of death in Switzerland with 37% of all deaths in 2004 (Junker, 2007). It is followed by cancer related deaths with 26%. While the percentage of cardiovascular related deaths of the 65 years and older increases to 41.7%, it still contributes with 20% of all deaths within the younger age group of the 45 to 64 year olds. Similar data exists in the EU with approximately 40% of all deaths related to cardiovascular diseases (Eurostat, 2002, Chapter 8).

1.1.2 Cardiac Disorders

In (Guidant, 2007) the following general types of heart diseases are identified:

Problems with the Heart's Electrical System. If the tissue that produces and sends electrical impulses to the heart muscle is damaged, abnormal heart beatings can occur in different variants: Bradycardia (unusually slow or unsteady rhythm), Tachycardia (more than 100 beats per minute), Atrial fibrillation (the upper chambers of the heart beat between 300 and 600 times per minute) and Sudden Cardiac Death (heart stops to beat and pump suddenly and unexpected).

Problems with the Heart's Arteries. These problems represent the most common heart disease in western countries. Fatty plaque builds up in the arteries and less blood flows through the heart muscle. This is illustrated in figure 1.

Problems with Arteries Outside the Heart. This disease type also identifies the building up of fatty plaque, but outside the heart, so less blood flows through the body. If the blocked artery is supplying the brain with blood, the patient may experience a stroke in the worst case.

A thorough introduction into cardiology is given in (Crawford et al., 2003).

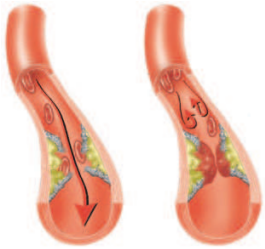


Figure 1: Coronary heart disease (Know Heart Diseases, 2007).

1.1.3 ECG

An electrocardiogram can be defined as a non-invasive test to reflect underlying heart conditions by measuring the electrical activity of the heart and producing an output in the form of a continuous strip graph. Early analysis and exploration of the heart's electrical system and its measuring techniques were examined by Willem Einthoven in the early 1900's (Biel et al., 2001).

The basic approach is to put electrodes on the body. The following three types of ECGs can be identified:

Resting. Taken within a couple of seconds.

Activity. Taken under physical activity.

Long-term. Usually taken during a 24 hour period.

The number of leads indicates the quality of the measurement. 12 leads ensure that all parts of the heart are recorded. It is generally accepted that ECG measurements with 3 leads ensure credibility (Burger and Osswald, 2003).

Various research has already been conducted regarding the transmission of ECG data. (Nollo et al., 2000) describes the national project "Territorial Tele-Cardiology" which focuses on providing experts with EGC data from medical ambulatory locations and patients homes. The average time of response is 1.5 days. The project demonstrated the feasibility and utility of a telecardiology network that offers improved access and quality in rural areas.

The design of a wearable device for ECG monitoring has been presented in (Led et al., 2004). The paper describes the usage of Bluetooth technology to transfer the ECG data to a mobile base unit.

Compression of ECG data is a topic that is very important in a homecare scenario, due to the limited bandwidth available over mobile networks and the high amount of data which is generated. An overview of various methods and research for the compression of biosignals is given in (Hadjileontiadis, 2006).

Table 1: Return rates.

Region		Infrastructure		Size of Location		
German	French	Hospital	Practice	$> 100'000$	$> 30'000$	$< 30'000$
35.7%	35.6%	61.5%	33.1%	29.2%	26.3%	45.4%

2 SURVEY RESULTS

The main motivation for the survey was to understand how cardiologists think about a telemonitoring system as an instrument for the detection of heart-related disorders. The main goal hereby was to research various components of how such a system should be built and what focus it should have in order to receive a high acceptance with the people that will eventually have to introduce such a system to the patients.

2.1 Approach

The printed questionnaires were sent on the 13th of March 2007 to 277 (out of totally 533) practicing cardiologists in Switzerland. French and German questionnaires were used to address the corresponding linguistic regions in Switzerland. The Italian region was not considered in this survey. 100 questionnaires were returned without any special follow ups, resulting in a total return rate of 36.1%. Table 1 illustrates the various return rates relative to various variables.

2.2 Structure of the Questionnaire

In order to design the questionnaire, some aspects from (Berekhoven et al., 2006) were applied:

- Interesting subject matter.
- Questionnaire as short as possible.
- Composition as interesting as possible.
- Questions as logical and easy as possible (optical as well).

Prior to the creation of the questionnaire an interview was held with a practicing cardiologist who has some background in the field of telecardiology. The result was a total of 12 questions that can be grouped into two categories:

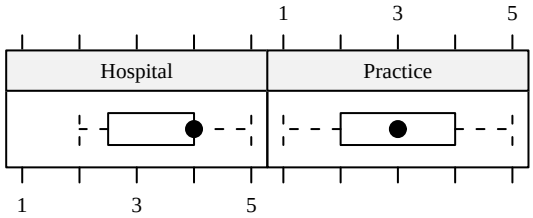
- Questions regarding ICT usage in cardiology in general.
- Specific mobile telecardiology application related questions.

A simplified process of a possible mobile telecardiology use case was illustrated on the first page of the questionnaire in order to give the respondents an idea about the intention of the survey.

The box-and-whisker diagrams used throughout this paper in order to illustrate the responses use a filled circle as median of the response and the position of the dot within the box gives an indication about the skew of the data. The dotted lines define the whiskers which can be seen as the smallest or biggest non-outlier observations. Eventual outliers are drawn as empty circles.

2.3 ICT in Cardiology

Two questions address the use of ICT within cardiology. In one question, the cardiologists were asked about their opinion on how they judge the desire of their patients to have innovative applications for treatment and diagnosis. 28.3% of the cardiolists think that their patients have a fairly strong desire for innovative applications. The correlations of these results with the demographic variables shows that the geographical region or the size of the location of the cardiologists does not influence the results. However, there is a correlation to the type of infrastructure the cardiologists operate in, indicating that cardiologists working in hospitals evaluate the desire of patients for new medical applications slightly higher than cardiologists in practices (see figure 2).

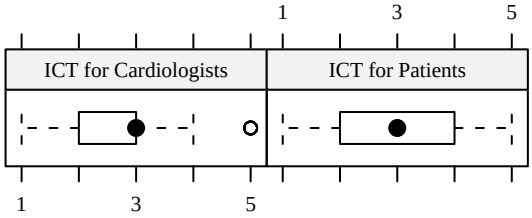


(1: strongly disagree, 5: strongly agree)

Figure 2: ICT usage for patients in cardiology.

The second question regarding ICT addresses the perspective of the cardiologists and asked if ICT is used sufficiently in cardiology. The mean is very similar to the former question. Compared to the infrastructure the cardiologists operate in, there is no difference between practices and hospitals, unlike to the former question. The results for both questions are illustrated in figure 3.

In summary, the following statements can be made in regard to ICT usage in cardiology:



(1: strongly disagree, 5: strongly agree)

Figure 3: ICT usage in cardiology.

- Patients have a certain desire for innovative applications in cardiology.
- Hospitalized patients have a higher desire for additional ICT in cardiology than patients in practices.
- Cardiologists feel that today's ICT usage in cardiology is fairly sufficient.

2.4 Motivation and Components in Telecardiology

Questions regarding the components in a telecardiology setting can be broken down into three main categories: The *Patient* category analyzes the abilities and the reasons of patients to use such a system. The questions in the category *Patient-end System Components* emphasize on various scenarios and number of leads that are necessary to measure an ECG. Furthermore, it investigates if the viewing of the ECG on a mobile base unit makes sense. The final category *Cardiologist-end System Components* ad

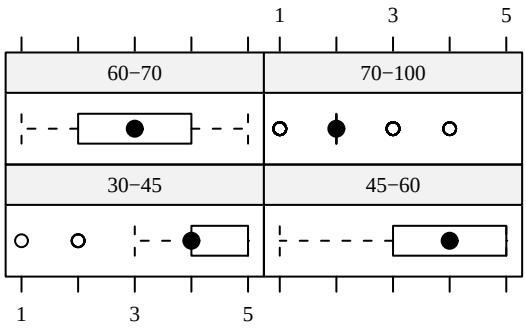
2.4.1 Patient Ability

The cardiologists were asked to evaluate the ability of their patients within certain age groups to record and send their ECG data in a given scenario. The results shown in figure 4 are not surprising. Cardiologists judge that about half of their 60 to 70 years old patients are able to record and send ECG data. The confidence is considerably higher for their younger patients.

2.4.2 Reasons for Patients

Possible reasons for using such a telecardiology system were investigated in a separate question. The total results, independent of the infrastructure or geographical region, are summarized in table 2.

Not surprisingly, the number one motivation is giving better access to patients living in rural areas. Cardiologists living in hospitals consider this even more important (75%) than their colleagues in practices (55%). The next four reasons are all rated with about the same importance. Interestingly, altough cardiologists assess elderly patients as less able to use such a mobile application (see figure 4), their limited mobility is ranked as number two reason. Furthermore, the possible time gain for patients using such a solution is judged positively by 50% of all cardiologists in hospitals as an important reason, whereas only 29% of practices see it like that.



(1: none, 3: about half, 5: almost all)

Figure 4: Patient ability per age group.

Table 2: Possible reasons for patients.

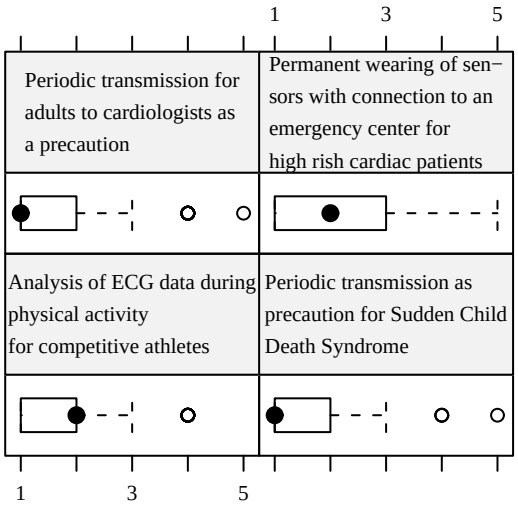
Reason	Agree
Cardiac patients who live in rural areas	58.6%
High quantity of elderly patients with limited mobility	42.4%
Higher service in quality and treatment	39.4%
Growth of mobile users with easiness to send ECG data	36.4%
Time gain for patients	32.3%
Reduction of costs	18.2%

2.4.3 System Components for Patients

This section of the questionnaire addresses various component settings and scenarios on the patient side for a mobile telecardiology application.

The first question considers various usage scenarios for patients and asked the cardiologists to evaluate their view which is illustrated in figure 5.

Only one of the proposed scenarios, the permanent transmission of ECG data for high risk patients, received a relatively high response. Regarding the periodic transmission of ECG data as a precaution, the cardiologists in big cities valued it slightly higher (mean 2.06) than their colleagues in towns (mean 1.79); this could be explained with the fact that cardiologists in big cities tend to treat more patients for routine check-ups.



(1: absolutely unnecessary, 5: absolutely necessary)

Figure 5: Usage scenarios.

The second question addresses the number of EGC leads necessary for a mobile application in order to reliably record the ECG of a patient. The distribution in figure 6 clearly illustrates two preferences: 3 leads and 12 leads.

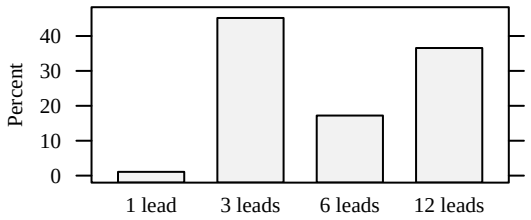
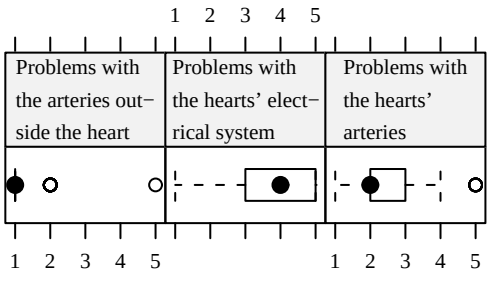


Figure 6: Number of leads preference.

The evaluation of the effectiveness of diagnostics using a mobile application for the three main types of cardiac diseases is addressed in the third question and its results illustrated in figure 7.

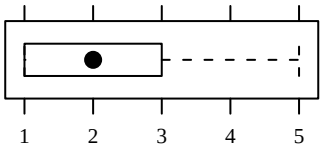


(1: not diagnosable, 5: very effective diagnosis)

Figure 7: Evaluation of disease types.

Problems with the arteries outside the heart do not represent situations that can be diagnosed using ECG sensor leads. In relation to the number of leads, patients that have problems with the hearts' arteries typically will use 12 leads and such applications will have to consider the larger data sets required. For patients that use such a mobile application as precaution for the heart's electrical system, a 3 lead ECG sensor appears to be more realistic.

The final question regarding the system components on the client side investigates the need of the patient to view his own ECG data on the mobile base unit. The results in figure 8 show that cardiologists do not consider such a functionality very desirable for their patients. This might be due to the fact that ECG diagrams can be quite complex to read and should thus only be analyzed by experienced professionals.



(1: no necessity, 5: strong necessity)

Figure 8: Patient desire to view ECG on MBU.

2.4.4 Components in Telecardiology for Cardiologists

The final part of this section considers the various components of a mobile telecardiology solution for the cardiologists themselves. It consists of two questions: The first question addresses the ICT knowledge of the cardiologists based on possible components or technologies that could be used for such a solution and is illustrated in figure 9.

Internet applications, mobile communication, information system and PDAs rank highest in the list and except for technical terms many possible components and technologies seem to be known reasonably well. Cardiologists in practices tend to have similar or slightly higher knowledge in regard to Internet Applications, as figure 10 illustrates. The same is true for mobile technologies and other fields.

This explains also the results of the next question: The cardiologists were asked about their preferences how they would like to receive the ECG data of the patients for the various scenarios already interrogated in figure 7. A summary of the preferences is ranked in table 3.

However, the results in figure 11 make it obvious that the preferred end system heavily relies on the kind of disease type that is handled with the solution.

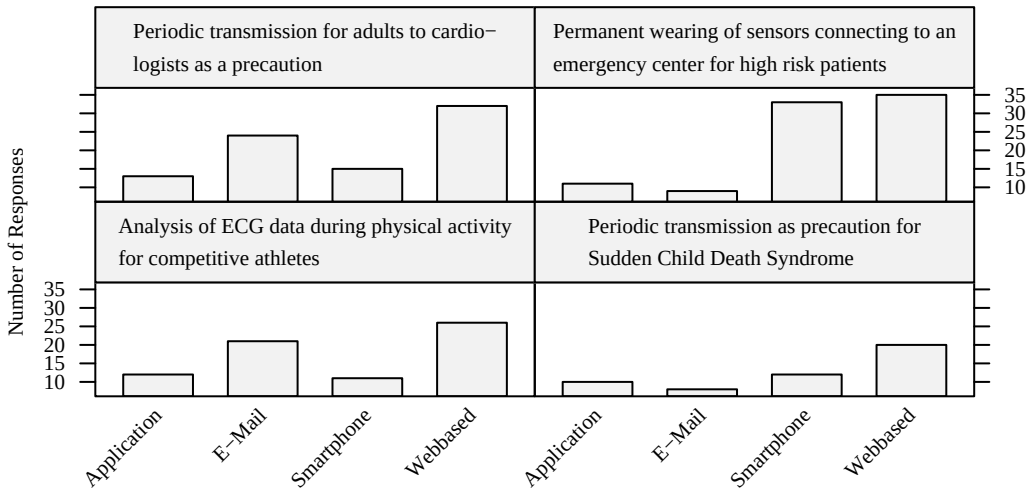
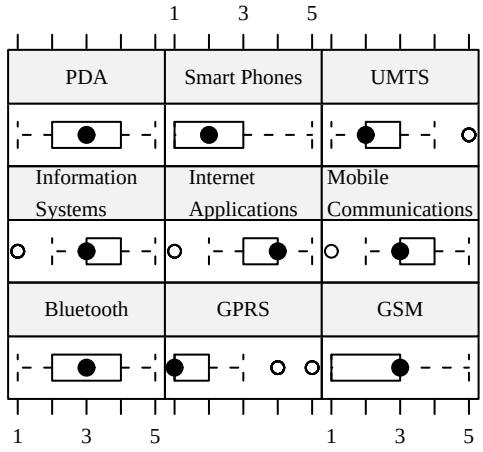
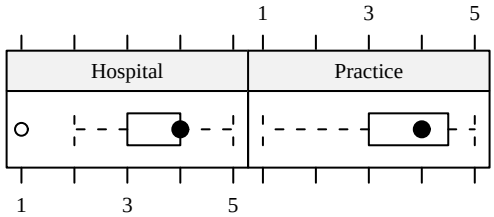


Figure 11: Preferred end systems per disease type.



(1: never heard of, 3: user knowledge, 5: extensive knowledge)

Figure 9: ICT knowledge of cardiologists.



(1: never heard of, 3: user knowledge, 5: extensive knowledge)

Figure 10: Knowledge about internet applications.

Web-based solutions seem to be a good choice in most cases. However, smart phones get a very high rating when the application handles high risk patients in an emergency. In such a situation, an e-mail that may take from minutes up to hours to arrive to the cardiologist, is not acceptable, even if the emergency is handled by a dedicated center. However, e-mails seems to be perfectly acceptable if the matter is about observing a patient over a longer period of time as a precaution.

Table 3: Ranked summary of preferred end systems.

System	Preference
Web-based solution	38.7%
Smart Phone	24.3%
E-Mail with attachment	21.2%
Dedicated application	15.8%

3 CONCLUSIONS AND OUTLOOK

The work in this paper focuses on the actual needs of a mobile telecardiology application with the results obtained from the survey. The conducted survey allows several conclusions for the implementation of a mobile application in cardiology:

- Except for the oldest age group (70-100), cardiologists agree that their patients are able to use such a system.
- Cardiologists do not seem enthusiastic about the proposed usage scenarios. However, a combination of factors, such as high risk patients in rural

areas, may strongly contribute to a higher acceptance.

- Addressing problems with the hearts' electrical system seems to be a widely accepted disease type for such a mobile application.
- There does not seem to be a strong desire for the patients to view their own EGC.
- The ICT knowledge of cardiologists and their preferences imply that, except for emergency situations, an internet application would be a good choice for their end system.

Security issues have explicitly been factored out of the questionnaire in order not to overwhelm the cardiologists with technical details. However, such issues are a pivotal part of every health related application and contribute highly to the acceptance and credibility of a system.

3.1 Outlook

In this paper we illustrated the perspective of cardiologists regarding a mobile telecardiology application scenario for patients. We would like to extend our research in the following areas:

- A survey on the patients may give insights about what elements to consider on the client side. It would be interesting to analyze the technological knowledge split by age group and the willingness of patients to use such a system when offered some well-defined benefits.
- The usage of standard equipment greatly contributes to a seamless rollout to the public. We would like to investigate how existing technologies can be integrated in order to provide a solution that aids the patients and cardiologists and that is also cost-effective.
- Qualitative interviews with cardiologists might give us a better idea about possible usage scenarios.

REFERENCES

Berekhoven, L., Eckert, W., and Ellenrieder, P. (2006). *Marktforschung – Methodische Grundlagen und praktische Anwendung*. Gabler Press.

Biel, L., Pettersson, O., Philipson, L., and Wide, P. (2001). ECG Analysis – A new approach in Human Identification. *IEEE Transactions on Instrumentation and Measurement*, 50:808–812.

Burger, P. and Osswald, S. (2003). Das EKG in der Telemedizin. *Swiss Medical Forum*, 3:661–643.

Crawford, M. H., DiMarco, J. P., and Paulus, W. J., editors (2003). *Cardiology*. Mosby, 2 edition.

del Pozo, F., de Toledo, P., Jimenez, S., Hernando, M. E., and Gomez, E. J. (2006). Chronic Patient's Management: The COPD example. In *M-Health – Emerging Mobile Health Systems*, pages 575–585. Springer.

Eurostat (2002). *Health statistics – Atlas on mortality in the European Union*. European Commission.

Fischer, H.-R., Reichlin, S., Gutzwiller, J.-P., Dyson, A., and Beglinger, C. (2006). Telemedicine as a new possibility to improve health care delivery. In *M-Health – Emerging Mobile Health Systems*, pages 203–218. Springer.

Guidant (2007). Heart Disease and Cardiovascular Disease Therapies. http://www.guidant.com. accessed on 5th June 2007.

Haas, P. (2006). eHealth verändert das Gesundheitswesen – Grundlagen, Anwendungen, Konsequenzen. *HMD Praxis der Wirtschaftsinformatik*, 251:6–19.

Hadjileontiadis, L. J. (2006). Biosignals and Compression Standards. In *M-Health – Emerging Mobile Health Systems*, pages 277–292. Springer.

Istepanian, R. S., Laxminarayan, S., and Pattichis, C. S. (2006). Ubiquitous m-health systems and the convergence towards 4G mobile technologies. In *M-Health – Emerging Mobile Health Systems*, pages 3–14. Springer.

Junker, C. (2007). *Todesursachenstatistik – Ursachen der Sterblichkeit 2003 und 2004*. Bundesamt für Statistik.

Know Heart Diseases (2007). Know Heart Diseases. http://www.know-heart-diseases.com. accessed on 5th June 2007.

Komninos, A. and Stamou, S. (2006). HealthPal: An Intelligent Personal Medical Assistant for Supporting the Self-Monitoring of Healthcare in the Ageing Society. In *Proceedings of UbiHealth 2006*.

Led, S., Fernandez, J., and Serrano, L. (2004). Design of a Wearable Device for ECG Continuous Monitoring Using Wireless Technology. In *Proceedings of the IEEE EMBS*, pages 3318–3321.

Leimeister, J. M., Krcmar, H., Horsch, A., and Kuhn, K. (2005). Mobile IT-Systeme im Gesundheitswesen, mobile Systeme für Patienten. *HMD – Praxis der Wirtschaftsinformatik*, 244:74–85.

Liszka, K. J., Mackin, M. A., Lichter, M. J., York, D., Pillai, D., and Rosenbaum, D. S. (2004). Keeping a Beat on the Heart. *IEEE Pervasive Computing*, 3(4):42–49.

Nollo, G., Nones, M., Malossi, C., Conforti, F., and Batocletti, B. (2000). Standard communication protocol for ECG exchange in tele-cardiology network for rural areas: three years project experience. *Computers in Cardiology*, 27:663–666.

Savini, M., Stormer, H., and Meier, A. (2007). Integrating Context Information in a Mobile Environment using the eSana Framework. In *Proceedings of the ECEH'07*, pages 131–142.

A HEART CELL GROUP MODEL FOR THE IDENTIFICATION OF MYOCARDIAL ISCHEMIA

Mohamed A. Mneimneh, Micheal T. Johnson and Richard J. Povinelli
Electrical and Computer Engineering, Marquette University, 1515 Wisconsin ave, Milwaukee, Wisconsin,USA
Mohamed.Mneimneh@Marquette.edu, Mike.Johnson@Marquette.edu, Richard.Povinelli@Marquette.edu

Keywords: Inverse problem, Ischemia, Decision Tree.

Abstract: Due to the increasing prices of medical care, and especially due to cardiovascular injury; scientists are looking for inexpensive and less invasive ways to diagnose myocardial ischemia. Many studies have shown that the variations of the ST-segment in the ECG signal are an indicator for ischemia. For this purpose, this work proposes an approach based on a heart cell group model and principle component analysis, using a decision tree classifier to differentiate between the ischemic and healthy beats. The cardiac based model is based on a physiological model of the electrical cycle of depolarization and repolarization of the atria and ventricles. The model parameters are estimated by minimizing the squared error between the generated signal and the recorded ECG. The approach is applied to beats from the Long-Term ST database, which consists of 86 subjects and more than 20,000 beats in which 80% of the beats are ischemic and 20% are healthy. A 10-fold cross validation test is performed over the dataset. The accuracy of this approach is 91.62%, with sensitivity of 95.09% and specificity of 75.66%.

1 INTRODUCTION

Ischemic heart disease is the leading cause of death in the world with almost 14% of all deaths (AHA 2005) Moreover, the average number of individuals who undergo a heart attack as a result of myocardial ischemia in the United States is approximately 1.5 million cases, of which 500,000 are fatal (AHA 2005). Myocardial ischemia is defined as the deprivation of oxygen in some portions of the cardiac tissue due to a blockage in the coronary artery. If the deprivation continues for an extended period, the effected cardiac tissue will die; thus, leading to a heart attack. Tissue that has died is no longer functional and diminishes the mechanical pumping function of the heart (Pardee 1920).

Early detection of ischemia is crucial because, in most cases, the effects of myocardial ischemia are reversible if detected early enough (Long 1980). General screening of patients is vital to preventing myocardial infarction, since ischemia can be present without exhibiting symptoms.

This work proposes a cardiac based model, Principle Component Analysis (PCA) and a C4.5 decision tree classifier for the detection of myocardial ischemia. The cardiac model is based on a physiological model of the electrical cycle of depolarization and repolarization of the atria and ventricles. The Sinoatrial (SA) node, the Atrioventricular (AV) node, bundle branches, Purkinje fibers, and left and right ventricles are modelled as signal generators. The ECG is generated by the difference in signal amplitudes arriving at the positive and negative terminals of an ECG lead. The model parameters are estimated through the minimization of the squared error between the generated signal and the recorded ECG. In addition to the obtained model parameters, 50 of the components from applying PCA to the signal are used in the diagnosis. A C4.5 decision tree is then used as a classifier to determine if a beat is healthy or ischemic.

The purpose of using electrocardiogram signals for the diagnoses of myocardial ischemia is because it is one of the least expensive techniques available to physicians. Figure 1 shows a labelled ECG signal showing the P, Q, R, S, T waves, the ST segment and the J point. The use of the ST level in the detection of myocardial ischemia was hypothesized in 1920 (Pardee 1920). Examples of low cost methods are ST event alerts ($250 cost) and easy to administer) with sensitivity of 46% and specificity of 91% and exercise stress testing ($200-$300 cost) with 68% accuracy of 68% (R. Gianrossi 1989).

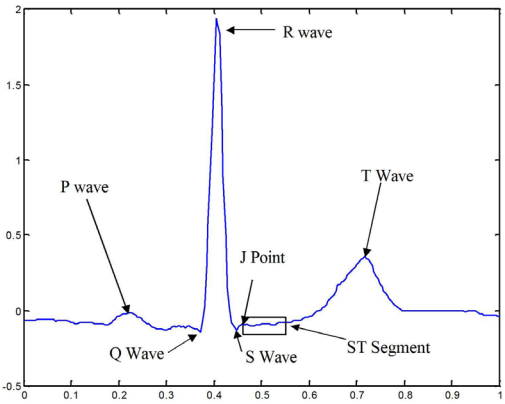


Figure 1: Labelled (ECG) signal.(Moody 2001).

Significant research has been undertaken to develop a more accurate, less invasive, and less expensive method for detecting myocardial ischemia. Much of this research focuses on the use of ECG signals. These methods build models or use thresholds of the ST deviation to determine if a patient's ECG signal might indicate ischemia.

Previous techniques that dealt with ischemia classification and detection when monitoring ECG signals started with low accuracy that increased significantly over time. These techniques are based on the hypothesis that myocardial ischemia can be detected by monitoring the ST variations.

Maglaveras et al. (N. Maglaveras 1994) have investigated a method for ischemia detection that uses supervised neural networks. The accuracy of this approach is of sensitivity of 73.0% and positive predictive accuracy of 69.5%.

RV Andreao et al. (R.V. Andreao 2004) employed a Hidden Markov Model for beat segmentation with the application of ischemia detection. The accuracy of this model is of sensitivity of 83.0% and positive predictive accuracy of 85%.

Additionally, T. Stamkopoulos et al. proposed an approach using nonlinear Principle Component Analysis (PCA) and neural networks in the identification of ischemic beats. The accuracy of this approach was 80% for healthy and 90% for ischemic beats when applied to the European ST-T Database (Stamkopoulos 1998).

Similarly, Victor-Emil Neagoe applied a Gaussian Neuro-Fuzzy Approach and PCA toward the classification of myocardial ischemia. The accuracy shown in the paper was 100% for 50 features. However, Neagoe dealt with only identifying ischemic and normal patients. Moreover, the number of training and testing data were 40 patients, half used for training and half for testing (Victor-Emil Neagoe 2003).

2 DATA SET AND PRE-PROCESSING

Various ECG and intracardiac datasets are available for the use of modelling and detecting myocardial ischemia. The data sets preserve the privacy of the subjects as there are no direct or indirect identifiers linking back to them.

2.1 Long Term ST Database

The Long-Term ST Database from PhysioNet contains 86 Holter ECG recordings from 80 independent patients. Holter recordings are ECG recordings recorded using portable recording devices, generally taken over a long period. These recordings were selected from the Holter libraries at Beth Israel Deaconess Medical Center in Boston, Physiolab (Laboratory of Biosignal Processing) of the Institute of Clinical Physiology in Pisa, Brigham and Womens Hospital in Boston, and the Zymed company. The recordings vary in length from 20 to 24 hours. Each record contains either two or three ECG leads. The records are digitized at 250 Hz with 12 bit resolution (Moody 2001).

Complete annotations have been provided for the database. These annotations label the significant ST shifts and episodes, the beginning (3-point) of most ST segments has been annotated along with R wave annotations using a 16 second averaging window. The beats were detected using WQRS function as part of the WFDB package supplied by the Physionet (Moody 2001).

To aid in the development of an ischemia classification algorithm, complete ST level annotations have also been provided. These annotations give the ST level, ST reference function, and the calculated ST deviation. The ST reference is expertly labelled moving average of the important ST shifts. The ST deviation is calculated by subtracting the ST level from the ST reference function shown in Figure 2 (Jager, Taddei & Moody 2003).

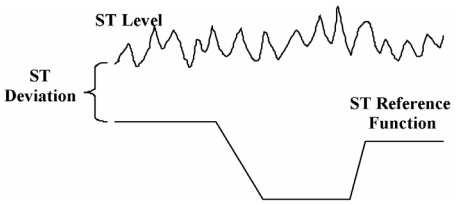


Figure 2: Example of ST deviation calculation.

The data consists of 43 free records from 42 patients and 43 fee records from 38 patients. The total number of beats used in this work is 20,528 for

both healthy and ischemic. The number of ischemic beats is 16,794, while that of the healthy beats is 3734.

In order to evaluate the proposed classifier, a ten fold cross validation is applied to the dataset. The ten fold cross validation is described as follows:

1. Divide data into 10 set of size n/10
2. Train on 9 sets and test on 1 set
3. Repeat the process 10 times and take the mean of the accuracy.

2.2 Signal Pre-processing

As mentioned in the previous section, the beats are obtained automatically from the records using the 'WQRS' function provided by the Physionet Toolkit. Each beat is then anchored such that the iso-electric line prior to the P wave is set to zero. A wavelet decomposition approach is used to denoise the signal from high frequency noise (GD. Clifford 2005).

3 METHOD

The classification approach utilizes a heart cell group model fitted to the patient's ECG signal along with the principle component analysis of the signal. The method is described in the block diagram shown in Figure 3.

A heart cell group model is used to generate a template ECG signal. Then, using a nonlinear constrained optimization technique, the model parameters are updated until reaching a certain error with the patient's signal beat. The estimated model fitting the ECG signal are then used with the PCA components as features in the C4.5 decision tree classifier.

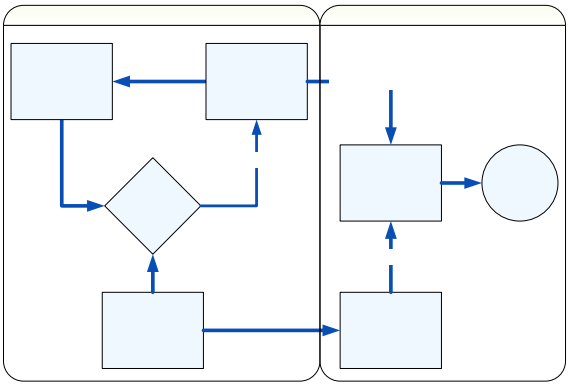


Figure 3: Block diagram of the Ischemia diagnosing method.

3.1 Heart Cell Group Model

Electrocardiograms indicate the electric activity of the heart over the body surface. In general, two types of model have been developed to characterize the ECG signal. The first type is a model used for interpolating experimental data and can be fitted to ECG signals without having a reference to the physical system. The second type is a model that can characterize the ECG signal and can be related back to the heart activity. The objective of this work is the latter modelling approach, focusing on development of a model that can estimate the activation sequences of the heart cells from real patient ECG signals. This objective is called the inverse problem. The difficulty of this problem is that unless it is stated in a particular manner, the solution will not be uniquely defined.

Several techniques have been employed for generating models to solve the forward and inverse problem. These techniques overcome the uniqueness problem by modelling the heart as a small number of moving dipoles. Some of these techniques apply the solution of Green's theorem (Method of Moments) or Multi-Pole technique to determine the scattering of the electric waves over the heart. These methods are considered accurate. However, the main drawback of these techniques is the computational complexity (Gulrajani 1998).

McSharry et al. presented a "dynamic ECG model" that incorporates the ECG features as a combination of Gaussian functions. Although this model is easy to build, it cannot be related to the heart cell activity (GD. Clifford 2005).

3.1.1 Proposed Cell Group Model

A Heart Cell Model (HCM) is proposed in this work based on the reconstruction of the ECG signal using a cell group model. This model accounts for the wave propagation of the SA node, the AV node, the bundle branches, Purkinje fibers, and left and right ventricles. We hypothesize that the electric activity of a heart cell group can be represented by a difference of two sigmoid functions.

The electric activity of the myocardial cells is caused by the variation of the positively and negatively charged ions of the cells. As presented by researchers (Andrew J. Pullan 2005), the electric activity of the cell is given in Figure 4 and it can be approximated by a difference of two sigmoid functions as shown in Figure 5.

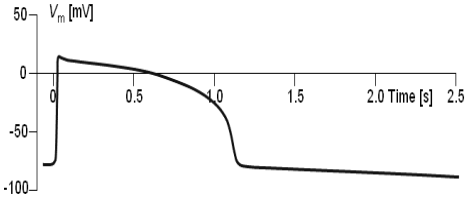


Figure 4: Conduction activity of the heart.

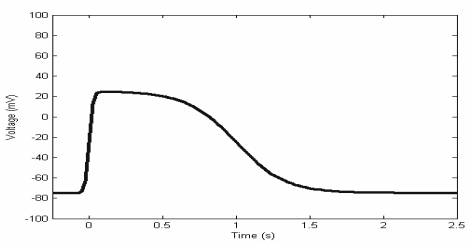


Figure 5: Proposed heart cell activity.

The cell group activity is modelled as the difference between two sigmoid functions:

$$f(t, a_1, c_1, a_2, c_2, k) = k\left(\frac{1}{1+e^{a_1(t-c_1)}} - \frac{1}{1+e^{a_2(t-c_2)}}\right), \quad (1)$$

where k represents the magnitude of the wave, a_1 and a_2 control the rising slope, and c_1 and c_2 control the translation in the direction of the *time* axis.

We hypothesize that the cumulative ECG signal is generated from the atrial and ventricular conduction activity. In this work, the P wave is assumed to be generated from the SA node activity; the PR interval from the AV node activity, and the QRS complex and T wave are generated from the activation of the bundle branches, the Purkinje fibers and, the right and left ventricles.

3.1.2 ECG Generation

As presented above, the ECG signal can be generated from the activation sequences of the heart cell groups. The same steps are used to generate the ECG from the modelled activation sequences. The model divides the heart into groups or nodes. Each node consists of a combination of cells at the SA node, the AV node, the bundle branches, the Purkinje fibers and, the right and left ventricles. Each node activation and deactivation sequence is represented as the difference between two sigmoid functions. The variables in the sigmoid functions consist of the magnitude, inflection (activation) point and the inclination slope. By summing the potential difference of the node signals at the positive and negative terminals of each lead, the ECG signal is generated:

$$\hat{f}_{ECG} = \sum_{i \in [SA, AV, Bb, Pf, LV, RV]} \left(f_i^+ - f_i^-\right), \quad (2)$$

where:

- SA and AV represent the activity of the SA and the AV node respectively.
- Bb and Pf represent the activity of the bundle branches and Purkinje fibers respectively
- LV and RV represent the activity of the Left and right ventricles respectively
- f^+ and f^- are the difference between two sigmoid functions as presented in (1) for each of the nodes at the positive and negative probes respectively

The following sections presents how the ECG wave features are generated. The features are the P wave, the PR segment, the Q wave, the R wave, and the S wave (QRS complex), ST segment, and T wave.

3.1.3 P Wave Generation

The P wave is generated from the potential difference between the electric conduction activity measured at the atrial cells at the positive and negative probes. In this approach, the atrial conduction activity at a single probe is estimated by equation (1). Moreover, it is hypothesized that the P wave can be generated from the conduction activity of the SA node:

$$P_{wave} = \left(f_{SA}^+ - f_{SA}^-\right), \quad (3)$$

The generation of the P wave using the difference of sigmoid estimation is shown in Figure 6.

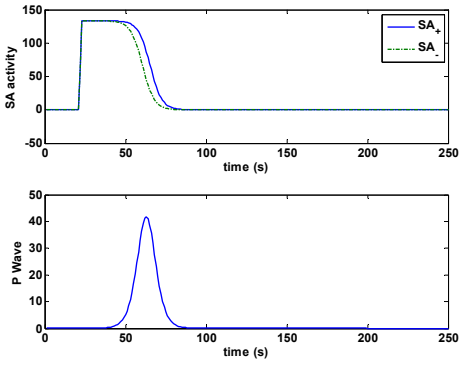


Figure 6: P wave generation using the differential sigmoid model.

3.1.4 PR Segment Generation

The PR segment occurs as the impulse travels from the AV node through the conducting tissue (bundle branches, and Purkinje fibers) towards the ventricles. Most of the delay in the PR segment occurs in the AV node. The PR segment is generally at the baseline; however, variations might occur due to certain heart diseases. Thus, by modelling the electric activity of the AV node as proposed in (1), and similar to the procedure shown in (3), we are able to generate the variations in the PR segment as shown in Figure 7.

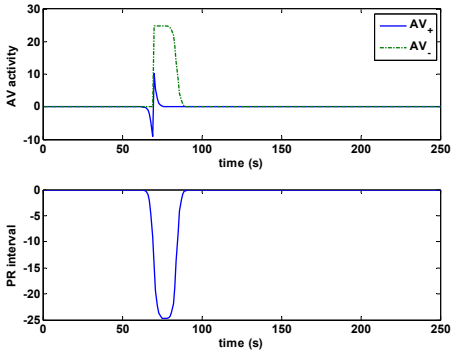


Figure 7: PR interval generation using the differential sigmoid model.

3.1.5 QRS Complex and T Wave Generation

The QRS complex and the T wave denote the interval for the beginning and end of the ventricular activation. When generating the QRS complex, the activity of the cell groups of the bundle branches, Purkinje fibers, and left and right ventricles are modelled during the ventricular cycle. The representation of the model for the QRS complex and T wave in an ECG signal is dependent on the difference between the positive and negative electrodes at the modelled cell groups. Figure 8 through Figure 10 show how each wave of the QRS complex and the T wave are generated using the differential sigmoid model.

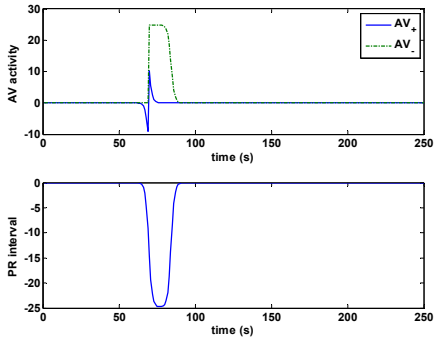


Figure 8: R wave and T wave generation.

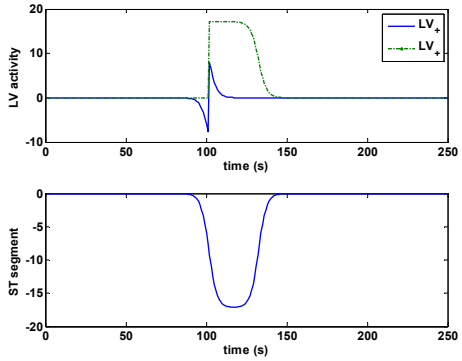


Figure 9: R wave and T wave generation.

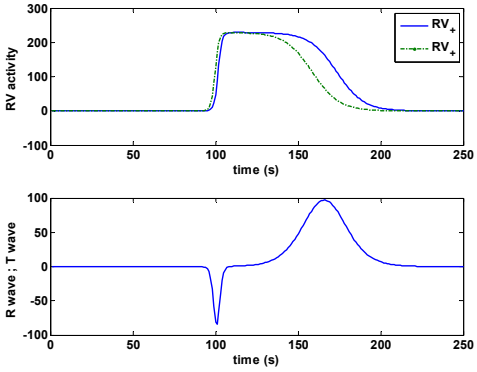


Figure 10: S wave and T wave generation.

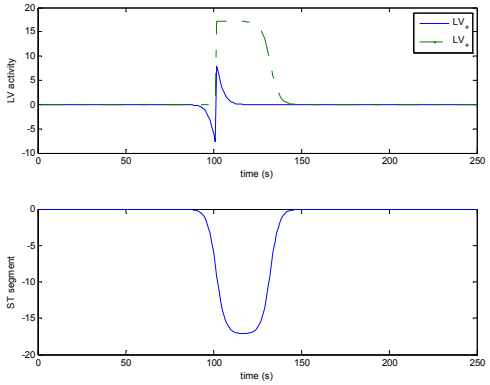


Figure 11: ST segment generation.

3.1.6 Parameter Estimation and Signal Fitting

This section discusses how to determine the parameters of the activation sequences in order to generate a real patient ECG signal. In order to achieve this task, a parameter estimation of the proposed model (1) and (2) is performed using the minimization of the least squares with the real ECG

signal. This process was performed with the help of the fmincon function, in Matlab, which finds a constrained minimum of a function for several variables. The function being minimized is given:

$$Error = \sum_{signal} \left(ECG - \hat{f}_{ECG} \right)^2, \quad (4)$$

The constraints applied to the function are that the atrial activity occurs prior to that of the ventricles. Moreover, the activation of the cell activity is constrained to occur prior to that of the deactivation. Additionally, the slopes of the activation are higher than those of the deactivation curves.

A template initial condition with known parameters for $\hat{f}_{ECG}$ is used to set the initial condition for the optimization process. Additionally, a dynamic template is generated for each beat. This choice of the template depends on the sign of the R peak. This allows more accuracy during the nonlinear optimization process. The highest cross-correlation point between the initial template the patient signal is then chosen.

Figure 12 shows the real and estimated ECG signal. It can be seen that the 'fitted' signal generated from the model matches the original patient signal. The model parameters used to generate the fitted signal are used as features in the classification process.

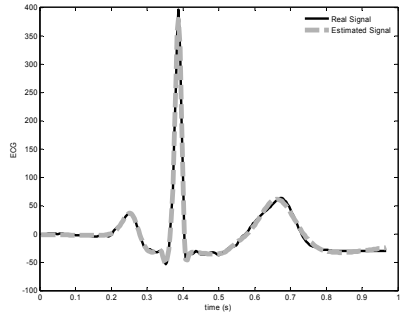


Figure 12: Estimated signal and original ECG signal.

3.2 Principle Component Analysis

Principle Component Analysis (PCA) is a linear transform where the basis functions are taken from the statistics of the signal, and can thus be adaptive. It is optimal in the sense of *energy compaction*, i.e it places as much energy as possible in as few coefficients as possible. The PCA is typically implemented using *Singular Value Decomposition*. The transform is generally not separable, and thus the full matrix multiplication must be performed:

$$X = U^T x, x = UX, \quad (5)$$

where the U is the basis for the transform. U is estimated from a number of x_i where $i \in [0 \cdots k]$:

$$\begin{aligned} U\Sigma V^T &= [x_1\, x_2 \ldots x_k] = A \\ U &= eigvec\left(AA^T \right) \end{aligned}. \quad (6)$$

3.3 C4.5 or J48 Decision Tree

Decision trees represent a supervised approach to classification. A decision tree is a simple structure where non-terminal nodes represent tests on one or more attributes and terminal nodes reflect decision outcomes. Generally, a decision tree algorithm chooses the attributes that best differentiates the output attribute values. The Weka classifier package (Eibe Frank 2007) has its own version of C4.5 known as J48. Weka's J48 is used in this work to solve the classification problem.

4 RESULTS

The HCM-PCA/C4.5 classifier is applied to the Long Term ST-Database. The proposed approach is compared to the technique proposed in (Stamkopoulos 1998). As mentioned before, the beat is detected using an automatic tool 'wqrs' provided by Physionet. The high frequency noise in the signal is removed using wavelet decomposition (Clifford 2006). The model is fitted to the model by minimizing the sum squared error using a constrained optimization process. The constraints are used to maintain the order of the heart's activation sequences. That is, the atrial activation occurs prior to that of the ventricles and the depolarization event occurs prior to the repolarization. The model parameters are used in the classification process, i.e. as features to determine whether a beat is ischemic or healthy.

A C4.5 decision tree is used in the classification process. As mentioned above, a 10 fold cross validation is performed. The classification method is applied with and without using the PCA components as features. Using the model parameters without the PCA features, the accuracy is 87.83% with sensitivity and specificity of 92.62% and 65.69%, respectively. Using the PCA features without the model parameters leads to an accuracy of 87.83% with sensitivity and specificity of 93.8% and 72.7%. However, when using the PCA features in addition to the model parameters, the accuracy increases to

91.62% with sensitivity of 94.89% and sensitivity of 75.66%. Sensitivity and specificity are defined as the accuracy of detecting the ischemic beat and the accuracy of detecting the non ischemic beat respectively. The confusion matrices for the proposed approaches are given in Table 1, Table 2, and Table 3 respectively. Confusion matrix is a visualization tool that presents the instances classified as ischemic or healthy in its columns and the actual classification in its rows.

Table 1: Confusion Matrix for HCM /C4.5 approach.

	Classified as	
	Ischemic	Healthy
Ischemic	15608	1255
Healthy	1243	2421

Table 2: Confusion Matrix for PCA/C4.5 approach.

	Classified as	
	Ischemic	Healthy
Ischemic	15877	986
Healthy	1044	2620

Table 3: Confusion Matrix for HCM-PCA/C4.5 approach.

	Classified as	
	Ischemic	Healthy
Ischemic	16035	828
Healthy	892	2772

It can be seen from Table 1 and Table 2 that the sensitivity of the proposed approach increases by 10% when using the PCA components in addition to the model parameters as features in the C4.5 decision tree classifier.

As mentioned above, the proposed approach is compared to the techniques of (Stamkopoulos 1998) as applied to the LT-ST database.

Table 4: Comparison between the proposed approach and previous methods.

Approach	Accuracy	Sensitivity	Specificity
HCM-PCA/C4.5	91.62%	94.89%	75.66%
Stamkopoulos	86.76%	91.73%	63.86%

It can be appreciated from Table 4 that the proposed HCM-PCA/C4.5 approach performs better than the previous methods by (Stamkopoulos 1998) for the LT-ST database. However, we have not been able to replicate the results of (Victor-Emil Neagoe 2003).

The importance in the proposed model, HCM, is that it can be related back to the heart's physical and electrical activity. It can be seen that the parameters of the HCM can be used in the detection of ischemic and healthy heart beats. This is due to the fact that the model parameters captured the information regarding the ECG waves and segments, such as slope, interval duration, magnitude and segment's variation.

5 CONCLUSIONS

A HCA-PCA/C4.5 approach is presented in this work to diagnose ischemic and healthy beats. The proposed approach is applied to the LT-ST database provided by Physionet. The approach showed excellent results when diagnosing ischemic and healthy beats. The proposed modelling approach provides a method to identify the features of ECG signals and an estimate to the cellular eclectic activity useful for ischemia detection. Finally, the proposed classification approach can be extended to detect different cardiac diseases.

REFERENCES

AHA 2005, *Heart Disease and Stroke Statistics*, Dallas, Texas: American Heart Association.

Andrew J. Pullan, M.L.B., Leo K. Cheng 2005, *Mathematically Modeling the Electrical Activity of the Heart from Cell to Body Surface and Back Again*, World Scientific, New Jersey.

Clifford, G. 2006, 'ECG Bag', <http://www.mit.edu/~gari/CODE/ECGtools/>.

Eibe Frank, M.H., Geoff Holmes, Mike Mayo, Bernhard Pfahringer, Tony Smith, Ian Witten 2007, *WEKA*, The University of Waikato.

GD. Clifford, P.M. 2005, 'Method to Filter ECGs and Evaluate Clinical Parameter Distortion Using Realistic ECG model Parameter Fitting', *Computers in Cardiology*.

Gulrajani, R.M. 1998, 'The forward and inverse problems of electrocardiography', vol. IEEE Engineering in Medicine and Biology Magazine, pp. 84 - 101.

Jager, F., Taddei, A. & Moody, G. 2003, 'Long-term ST database: A reference for the development and evaluation of automated ischaemia detectors and for the study of the dynamics of myocardial ischaemia.' *Medical and Biological Engineering and Computing*, vol. 41, pp. 172-171 182.

Long, C. 1980, *Prevention and rehabilitation in ischemic heart disease*, Baltimore: Williams & Wilkins.

Moody, J.M. 2001, *Preventive cardiology : strategies for the prevention and treatment of coronary artery disease*, Humana Press, Totowa, N.J.

N. Maglaveras, T.S., C. Pappas, and M. Strintzis 1994, 'Use of neural networks in detection of ischemic

episodes from ECG leads', *Neural Networks for Signal Processing*, pp. 518-524.
Pardee, H. 1920, 'An electrocardiographic sign of coronary artery obstruction', *Arch Int Med*, vol. 26, pp. 244-257.
R. Gianrossi, R.D., D. Mulvihill, K. Lehmann, P. Dubach, A. Colombo, D. McArthur, and V. Froelicher 1989, 'Exercise- induced ST depression in the diagnosis of coronary artery disease. A meta-analysis', *Circulation*, vol. 80, pp. 87-98.
R.V. Andreao, B.D., J. Boudy, J.C.M. Mota 2004, 'ST-segment analysis using hidden Markov Model beat segmentation: application to ischemia detection', *Computers in Cardiology*, pp. 381 - 384.
Stamkopoulos, T.D., K. Maglaveras, N. Strintzis, M. 1998, 'ECG analysis using nonlinear PCA neural networks for ischemia detection', *IEEE Transactions on Signal Processing*, vol. 46, no. 11.
Victor-Emil Neagoe, I.-F.I., Sorin Grunwald 2003, 'A Neuro-Fuzzy Approach to Classification of ECG Signals for Ischemic Heart Disease Diagnosis', *AMIA Annu Symp Proc*, p. 494–498.

APPENDIX

The cost function for the constrained optimization function is obtained by replacing (1) into (2):

$$\min_{x} \sum \left(ECG - \sum_{\substack{i=SA,AV,Bb,\\ Pf,LV,RV}} k_i \cdot g\left(t, \mathbf{a}_i, \mathbf{c}_i\right) \right)^2 , \qquad (7)$$

$$g\left(t, \mathbf{a}_i, \mathbf{c}_i\right) = d\left(t, a_{i,1}, c_{i,1}, a_{i,2}, c_{i,2}\right) - d\left(t, a_{i,3}, c_{i,3}, a_{i,4}, c_{i,4}\right) , \qquad (8)$$

$$d\left(t, a_1, c_1, a_2, c_2\right) = s\left(t, a_1, c_1\right) - s\left(t, a_2, c_2\right) , \qquad (9)$$

$$s\left(t, a, c\right) = \frac{1}{1 + e^{-a(t-c)}} , \qquad (10)$$

Subject to the constraints: $c_{(i)1,3} < c_{(i)2,4}$, $c_{(SA)1,3} < c_{(AV)1,3}$, $c_{(SA)1,3} < c_{(Rvep)1,3}$, $c_{(SA)2,4} < c_{(AV)1,3}$, $c_{(AV)2,4} < c_{(Rvep)1,3}$, $c_{(AV)2,4} < c_{(Rven)1,3}$, $c_{(AV)2,4} < c_{(Lvep)1,3}$, $c_{(AV)2,4} < c_{(Lven)1,3}$, $c_{(Rven)2,4} < c_{(Lvep)1,3}$, $c_{(Rven)2,4} < c_{(Rvep)1,3}$, $c_{(Rvep)1,3} < c_{(Lvep)1,3}$, $c_{(Lvep)2,4} < c_{(Rvep)2,4}$, $c_{(Rvep)1,3} < c_{(Lven)1,3}$, $c_{(Lven)2,4} < c_{(Rvep)2,4}$, $c_{(Lvep)1,3} < c_{(Lven)1,3}$, $c_{(Lven)2,4} < c_{(Lvep)2,4}$,

$0.05 < \left|c_{(SA)4} - c_{(SA)2}\right| < 0.12$, $0.05 < \left|c_{(SA)4} - c_{(SA)2}\right| < 0.12$

$0.05 < \left|c_{(AV)4} - c_{(AV)2}\right| < 0.10$, $0.05 < \left|c_{(Rvep)3} - c_{(Rvep)1}\right| < 0.08$,

$0.05 < \left|c_{(Rvep)4} - c_{(Rvep)2}\right| < 0.10$, $0.05 < \left|c_{(Rven)4} - c_{(Rven)2}\right| < 0.03$,

$0.05 < \left|c_{(Lvep)3} - c_{(Lvep)1}\right| < 0.03$, $0.05 < \left|c_{(Lvep)4} - c_{(Lvep)2}\right| < 0.10$,

$0.05 < \left|c_{(Lvep)3} - c_{(Lvep)4}\right| < 0.10$.

SEGMENTATION AND CLASSIFICATION OF CUTANEOUS ULCERS IN DIGITAL IMAGES THROUGH ARTIFICIAL NEURAL NETWORKS

André de Souza Tarallo, Adilson Gonzaga
Department of Electric Engineering, USP University, Av.Trabalhador São Carlense- 400, São Carlos, Brazil
tarallo@sel.eesc.usp.br, adilson@sel.eesc.usp.br

Marco Andrey Cipriano Frade
Department of Medical Clinics, USP University, Ribeirão Preto, Brazil
mandrey@fmrp.usp.br

Keywords: Leg Ulcer, Computer Vision, Artificial Neural Network.

Abstract: Treatments of leg ulcers are generally expensive and those conducted through the direct manipulation for analysis of its evolution. The treatment efficiency is observed through the reduction of the size of ulcers in relation to the amount of tissues found in their beds, which are classified as granulated/slough. These results are obtained through analyses performed after consultation due to the time these analyses take. This work proposes a new non-invasive technique for the follow-up of treatments aimed at cutaneous ulcers. In this methodology, it was proposed that digital photos of cutaneous ulcers would be submitted to an artificial neural network (ANN), so that all surrounding the wound except for the wound itself could be extracted (skin/background), thus obtaining the ulcerated area. Computer vision techniques have been applied in order to classify the different types of tissues found in the ulcer bed, thus obtaining the corresponding granulation and slough percentages as well as its area. The results obtained have been compared with the results obtained by Image J software. Finally, this methodology will be a useful tool for health professionals in relation to the quickness and precision that it will provide results along the consultation.

1 INTRODUCTION

Leg ulcers are a public health problem worldwide and reach from 3% to 5% of the population older than 65 years of age and 1% of the adult population. (Mekkes et al., 2003). The treatment presents some complications due to its long-term characteristic, discomfort of curatives and uncertainness in relation to its success, once its cure depends on several factors that act as intervenient variables in the process, causing significant social and economic impact. The treatment is painful, expensive and slow due to a number of associated etiopathogenic factors, and the disease represents one of the main causes for work absenteeism.

The use of computer tools involving image processing (Computer Vision) and ANN consists of an alternative analysis method for the follow-up of leg ulcer treatments. (Albu and Ungureanu, 2005). This method does not allow the direct contact with the wound, once ulcers are analyzed through digitized images (Goldman and Salcido, 2002). Therefore, the health professional disposes of tool designed to support the treatment of ulcers.

The objective of this work is to present a proposal to aid in the quantitative analysis of each tissue found in the inner part of wounds, which are classified as granulated and slough and in the calculation of the wounded area. With these measurements, one may have a perspective in relation to the treatment evolution, since it provides a dynamic-therapeutic healing follow-up.

This work also proposes the development of methodology to classify leg ulcer tissues in order to support specialists along the treatment evolution. The employment of computer software with the proposed methodology may lead the patient to feel safer, since there is no direct contact with the wound to obtain samples for analyses. In a first phase, the proposal consists of performing the extraction of features from the leg ulcer digital image base through color samples removed from ulcer images manually and of applying them to the neural

network test for images segmentation – Training Phase. In a second phase image processing techniques were used to classify tissues found at the inner region of the wound - Test Phase.

2 LEG ULCERS

Leg ulcers occur due to more than one cause: venous hypertension (~80%), arterial insufficiency (~10%) or the combination of both – the called "mixed ulcers", among others. (Abbade and Lastória, 2005). Ulcers occurring at the tip end of the lower limbs are a result of venous, arterial or neurovascular diseases such as varix, thrombosis venous, arteriosclerosis, diabetes and arterial hypertension, among others. (Dean, 2006).

Each type of ulcer presents own characteristics, requires different treatments, and must be evaluated separately. (Borges, 2005 and Dean, 2006).

A deficient blood circulation decreases the intake of oxygen and nutrients and reduces the removal of metabolism-derived products such as free radicals, factors that contribute for the healing delay. The main problem of leg ulcers is the recurrence; 30% of healed ulcers recur within the first year and this rate increases to 78% after two years when inadequately treated. (Barros, 2000).

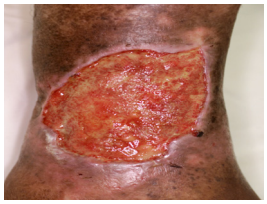


Figure 1: Venous Leg Ulcer.

A good analysis of some characteristics, parameters and interpretation of clinical ulcer examinations is vital, and among these characteristics: number and size of the ulcer, edges and appearance of the lesion bottom, type of ulcer, skin-ulcer state, arterial test, venous test and evaluation of the microbiological status (culture, exams) are worth mentioning. (Kupcinskas, 2006).

Leg ulcer is a very relevant and common problem in health services worldwide and affects between 0.1% and 1% of the adult population with studies pointing to prevalence. (Mekkes et al., 2003). However, the key for the selection of effective ulcer treatments is based on the evaluation process of its etiology.

Venous-origin leg ulcers – Figure 1 – popularly known as varicose ulcers, are mainly caused by chronic venous insufficiency, term described as lower limbs syndrome, which represents the incapacity of maintaining the balance between the arterial blood flow that reaches the lower limb and the venous flow that returns to the right atrium as a result of the incompetence of the superficial and/or deep venous system with symptoms such as edema, pigmentation, pain and disabilities. (Barros, 2000) (Pitta, Castro and Burihan, 2000).

In normal people, the blood pressure decreases during the practice of physical exercises, but in patients with venous incompetence, the pressure remains high during effort. Venous ulcers are mainly characterized by the presence of edema, darkened pigmentation, varicose veins and lipodermatosclerosis (hardening and fibrosis in the dermis and subcutaneous tissue) at the lower limbs. (Phillips and Dover, 1991).

In the study conducted by Skaraborg (cited in Figueiredo, 2003), 5.6% of people with 65 years of age or older presented open or healed lower limbs ulceration and 2.4% of the adult population above 15 years of age have already had ulcers. European data show that 1.5% of the adults will have ecstasy ulcer sometime in their lives.

3 MATERIAL AND METHODS

The photographs were taken through a Sony Cyber shot P-93 camera with 3 mega pixels, 3X optical zoom and without digital zoom. The images randomly selected from our the image bank were standardized and non-standardized in relation to zoom, illumination, distance between the camera and the patient's leg and the focus in the patient's leg. We made the image bank because none was

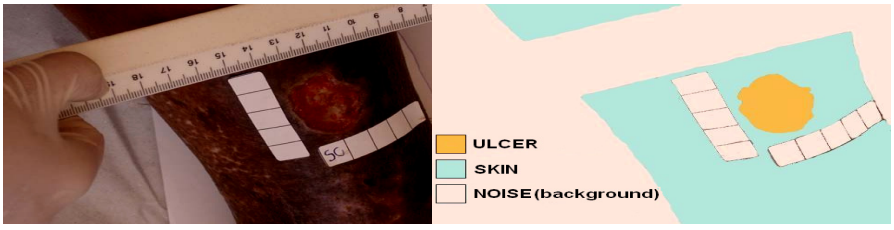


Figure 2: Example of an image with noise, skin and ulcer regions.

found publicly available library. Fifty images of thirty five patients were selected to test the validity of the proposed methodology.

The methodology proposed is divided into two phases: in the first phase, the extraction of the color characteristic and the ANN training occur (Training Phase) with ten images. (Haykin, 2001). The second phase consists of segmenting images with forty images (ANN Test), elimination of noises, improvement of the image quality and later tissue classification in the wound bed – Test Phase. (Gonzalez and Woods, 2002).

In the first phase, initially two algorithms were applied to images in order to obtain, skin, ulcer (bed) and noise (background - all that is not skin and ulcerated area) color - Figure 2-, which will serve as inputs for the ANN training to distinguish the color characteristics of the wound edge from the other colors not involved in the wound thus, forming training standards.

The color characteristics corresponding to skin and non-skin (skin/noise/background) in the RGB model are obtained through the first algorithm; this process is manually performed by the computer operator (this process should be performed by a health professional, once he will know which are the best points to be selected in order to find out what each color represents in the image). The software used for the development of this methodology was the Matlab 7.0, (Math Works, 2004) which shows the 50 images selected (one at a time) and waits for the computer operator to select the image region with the mouse with the aid of the algorithm. Each color characteristic of the selected region is stored in a text-type file to form the feature vector (skin/non-skin matrix), according to Figure 3.

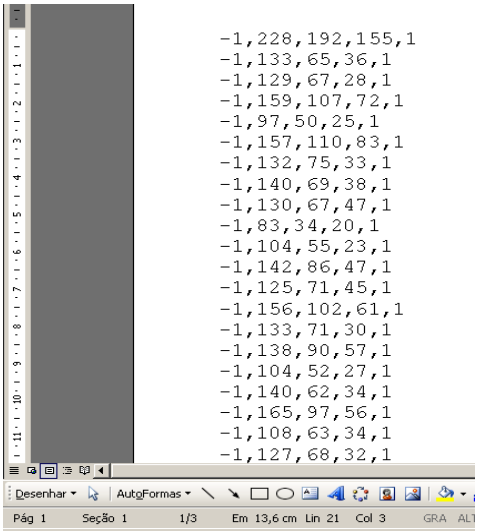


Figure 3: Example of Skin/Non-Skin Matrix.

If one photo contains several interesting characteristic regions, this image is opened more than once for the selection of the characteristics.

The values presented in each line of Figure 3 represent the following:

- -1 is a bias used by the neural network for the activation of the neuron;
- The three next values refer to the RGB value in relation to the color selected by the user;
- 1 is the value to be used as exit desired by the neural network.

The second algorithm is used to obtain the wound color characteristics in the RGB model, which are obtained as in the first algorithm. The feature vector (wound matrix) of each selected color is saved in another text-type file. The desired exit of the wound matrix is the 1.

These two matrixes will form the "training patterns", which will be used for the training.

The first phase of the proposed methodology in divided into two stages. In the first stage, the entrance characteristics for the neural network are obtained (color characteristics) and in the second one, these characteristics are applied in the neural network for its training – Figure 4.

3.1 ANN Training

The extracted characteristics (training patterns) are applied to an ANN for its training and later classification and separation of the wound from the remaining portion of the image (Test Phase). The MLP Feedforward neural network architecture was used with the Back-propagation training algorithm (Haykin, 2001), which was the architecture most used for classification in several areas, and the cutaneous ulcer images were generated in the RGB color model. Before the Test Phase, the ANN must be trained in order to learn about the color characteristics obtained through both algorithms previously mentioned. The training characteristics are the following:

- Both features vectors are concatenated in order to form the training matrix. Bias, RGB and the desirable output are arranged in different variables, and the RGB characteristics are normalized for the [-1, 1] interval.
- The neural network is initialized using the minimum/maximum function of the training matrix.

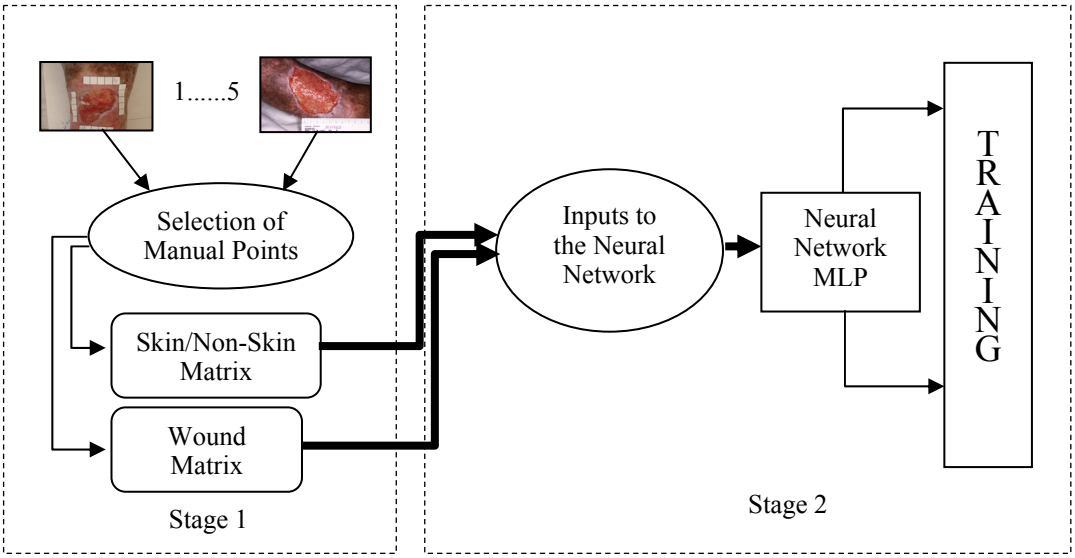


Figure 4: First phase of the proposed methodology.

- The neural network training was performed using the tangent-hyperbolic sigmoid activation function (so that the values corresponding to the RGB characteristics do not exceed the normalized interval). The moment gradient is used for the three occult layers of the neural network plus the output layer.
- Values corresponding to other parameters used in this algorithm and in the Neural Network will be specified in the next topics.

3.2 ANN Test (Classification of Images)

In the second phase, or Test Phase, the efficiency of the Neural Network is verified in the segmentation of the 40 images from results obtained in the training (first phase). A post-processing is required to eliminate some remaining noises to better prepare the image for the tissue classification. Figure 5 presents the second phase of the proposed methodology. The techniques employed are the erosion and dilation morphologic operations. (Gonzalez and Woods, 2002). Finally, the tissues are classified based on the counting of pixels, where similar colors are associated to the type of tissue. Besides the granulation and slough tissue classification, the percentage of these two types of tissues and the ulcer area in the image were calculated.

The algorithm used (ANN Test) presents the following steps:

1) Segmentation of Images;
2) Post-Processing: Images are processed through dilation and erosion and image superposition in order to eliminate noises and to show the wound region only;
3) Counting of pixels corresponding to the granulation and slough tissue and calculation of the percentage corresponding to each type of tissue in the image;
4) Generation of an image with markings in which pixels corresponding to granulation and slough are counted: white pixels are granulation tissue and the others are the slough tissue;
5) Calculation of the leg ulcer wounded area in cm2.

The segmentation is performed by the Neural Network using parameters from the Training Set and commands based on Neural Networks toolbox from the Matlab 7.0 software (Math Works, 2004); and the resulting image that distinguishes the wound from the rest of the image is obtained – pre-processed image.

3.2.1 Post-Processing

The pre-processed image is then submitted to a post-processing in order to eliminate noises and to show the wound region only. To do so, erosion and dilation morphologic operators were used.

In order to use erosion and dilation operators in the Matlab software, the figure has to be converted into gray scale, which is the only way that the figure allows the use of such morphologic operators (Gonzalez and Woods, 2002).

In order to use these morphologic operators, a structuring element should be created to serve a

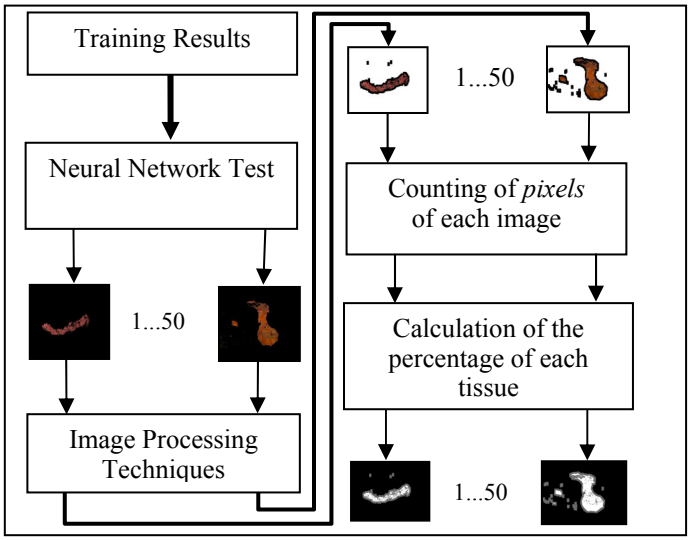


Figure 5: Second phase of the proposed methodology.

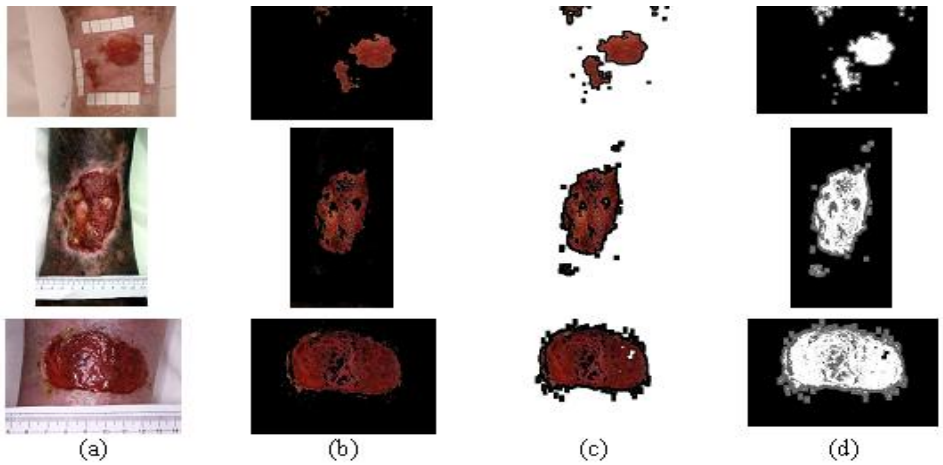


Figure 6: Leg Ulcer Segmentation Results. In (a) Original Image – (b) Pre-Processed Image – (c) Post-Processed Image – (d) Image with counted pixels.

dilation parameter. The structuring element used had a square format (Math Works, 2004). Following, the Sobel edge detector was used.

This image was superposed to the original image with the objective of obtaining an improved and less noisy new image – post-processed image in which pixels were counted and calculations were performed. Finally, an image based on the post-processed image was generated, with marking of the sites in which slough and granulation pixels were counted. Figure 6 presents images according to the algorithm execution sequence. The parameters used in the neural network of this methodology may be observed in Table 1.

Table 1: Values of the Neural Network Parameters.

Parameter	Value
Neurons in the 1st hidden layer	4
Neurons in the 2nd hidden layer	4
Neurons in the 3rd hidden layer	1
Moment Term	0.5
Maximum Number of Iterations	1000
Training Error Rate	$1x10^{-3}$

4 RESULTS AND ANALYSES

Considering the 40 test images and the segmented wound area only, the average slough and granulation

percentages in relation to the total image may be verified in Table 2. The results were obtained through the proposed methodology.

Table 2: Arithmetic average of the tissue percentage– Proposed Methodology.

	Total Image	Wound Area
Slough	10.5%	26.1%
Granulation	18.4%	73.9%

The same images tested in the proposed methodology were applied to the Image J for comparison purposes, because this software is used for made analyses of images of leg ulcers at department of dermatology of FMRP (Ribeirão Preto Medical School) and it is desirable to have a tool more practical than the Image J.

The results obtained through Image J may be observed in Table 3. (Gomes, Santana and Minatel, 2005).

Table 3: Arithmetic average of the tissue percentage – Image J.

	Total Image	Wound Area
Slough	18.9%	43.3%
Granulation	30.0%	56.7%

The area of each wound in cm^2 in relation to the total image was also calculated both through the proposed methodology and through the Image J; the arithmetic averages of results may be observed in Table 4.

Table 4: Arithmetic average of the Wound Areas.

	Proposed Methodology	Image J
Average Area	13.1cm^2	14.1cm^2

The results obtained through the Image J freeware software and with our methodology seemed to be satisfactory; in the total tissue area, the average was 13.1 cm^2 through the proposed methodology and 14.1 cm^2 through Image J (Figure 7). In relation to the granulation, the average obtained was 12.4 cm^2 through the proposed methodology and 12.6 cm^2 through Image J (Figure 8). In relation to slough, the average obtained was 1.8 cm^2 through and 1.9 cm^2 through Image J (Figure 9).

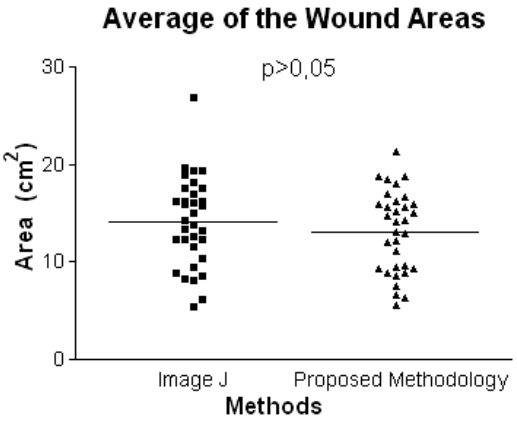


Figure 7: Results of the t Test for total area.

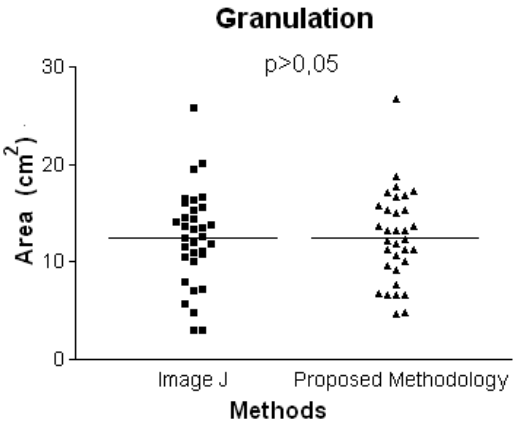


Figure 8: Results of the t Test for granulation.

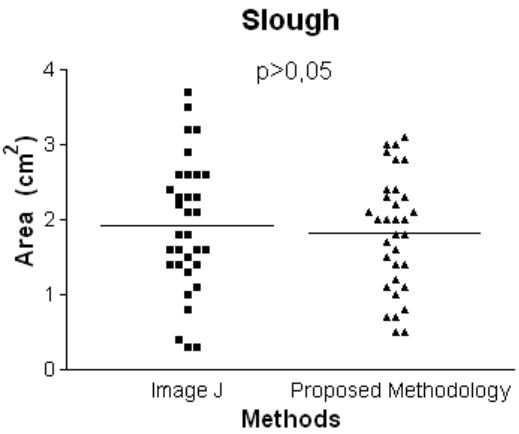


Figure 9: Results of the t Test for slough.

It is worth reminding that the area evidenced through Image J is manually performed, and takes a long time until it comes to the final results, whereas in the proposed methodology, this process is automatically performed by the neural network,

which makes the processing faster and safer. This area evidenced manual affects in the difference of the results of the table 2 and 3 as well as some terrible interpretations of RNA owed the qualities of the images and of obtaining of the same ones.

The results were analyzed by a medical area specialist, who verified the concordance of results obtained.

Figure 7 shows the graphic of the t-student test applied to results obtained through both the proposed methodology and Image J for total areas. Lines in the center of the graphic show the arithmetic averages of results obtained through each methodology and one may observe that they are very close to each other.

Similarly, there are two other graphics that also corroborate the efficiency of results obtained through both Image J and the approach of this paper. Figure 8 shows the results of t-student tests for granulation area and Figure 9 for slough area.

5 CONCLUSIONS

Both the Image J and our methodology based on ANN presented satisfactory results. The t-student test at 95% was applied and the results confirmed the efficiency of both methods. This finding testifies that the variation observed between the results obtained through both methodologies is acceptable and that they can be applied in practice.

The results obtained suggest that both image analysis methods are effective in the measurement of total area, granulation and slough, being considered as adequate for the dynamic-therapeutic evaluation of leg ulcers. Artificial Neural Networks seem to be a high-level methodology for the analysis of images due to the lower interference from the operator/researcher, since it does not require manual design.

This new application will be one more tool to aid in the diagnosis at FMRP and perhaps replace the image J because of its little practicality. For better performance of this new application is desirable to use standardized images, as mentioned in item 3, because the images non-standardized not behaved so well on the standardized; but nevertheless been achieved good and acceptable results general finals.

This project encourages and contributes for the application of new technologies and hence the use of softwares in this area with the emergence of new research lines.

REFERENCES

Abbade, L. P. F., Lastória, S., 2005. Venous ulcer: epidemiology, physiopathology, diagnosis and treatment. *International Journal of Dermatology*, Vol. 44, pp. 449 – 456.

Albu, A., Ungureanu, L., 2005. Artificial Neural Network in Medicine. In *Symposium on Applied Computational Intelligence*.

Barros, J. R., 2000. Insuficiência Venosa Crônica. In Pitta, G. B. B., Castro, A.A. & Burihan, E. Angiologia e Cirurgia Vascular: Guia Ilustrado, in *UNISAL/ECMAL,* Brazil, Maceió.

Borges, E. L., 2005. Tratamento Tópico de Úlcera Venosa: Proposta de Uma Diretriz Baseada em Evidência, MSc. Thesis, University of São Paulo.

Dean, S., 2006. Leg Ulcers – Causes e Management. *Reprinted from Australian Family Physician,* Vol. 35, no. 7, pp. 480-484.

Figueiredo, M., 2003. Úlcera Varicosa. Angiologia e Cirurgia Vascular: Guia Ilustrado, viewed, 13 April 2007, <*http://www.lava.med.br/livro/pdf/marcondes_ulcera.pdf*>.

Goldman, R.J., Salcido, R., 2002. More Than One Way to Measure a Wound: An Overview of Tools and Techniques – Clinical Management. *Advances in Skin & Wound Care*, vol.15, no.5.

Gomes, F.G., Santana, L. A. & Minatel, D. G., 2005. Uso do Software Image J Para Análise Clínico-Fotográfica das Úlceras de Perna. *5º. Encobio "Encontro de Bioengenharia"*, University of São Paulo, pp. 37.

Gonzalez, R., Woods, R., 2002, *Digital Image Processing*, 2nd edn, Prentice Hall.

Haykin, S., 2001, *Neural Networks: A Comprehensive Foundation*, 2nd edn, Prentice Hall.

Kupcinskas, A. J., 2006, viewed 13 February 2006, <http://www.ajkj.med.br/ulc.htm>.

Mekkes, J.R., Loots, M.A.M., Van Der Wal, A.C., & Bos, J.D., 2003. Causes, Investigation And Treatment Of Leg Ulceration. *British Journal of Dermatology*, vol.148, pp. 388-401.

Phillips, T. J., Dover, J. S., 1991. Leg Ulcers. *J Am Acad Dermatol*, St Louis, vol.25, pp. 965-987.

Pitta, G. B. B., Castro, A. A. & Burihan, E, 2000. Angiologia e Cirurgia Vascular: Guia Ilustrado, in *UNISAL/ECMAL,* Brazil, Maceió.

Math Works, 2004, ver. 7, computer program, The Math Works Inc., Matlab Help, USA.

MEDICAL IMAGE MINING ON THE BASE OF DESCRIPTIVE IMAGE ALGEBRAS
Cytological Specimen Case

I. Gurevich, V. Yashina
Dorodnicyn Computing Center of the Russian Academy of Sciences, Vavilov str., Moscow, Russian Federation
igourevi@ccas.ru, lcmi@ccas.ru, werayashina@mail.ru

H. Niemann
University of Erlangen-Nuernberg, Lehrstuhl fuer Informatik, Martensstr., Erlangen, Germany
niemann@informatik.uni-erlangen.de

O. Salvetti
Institute of Information Science and Technologies, CNR, 1, Via G.Moruzzi, Piza, 56124, Italy
Ovidio.Salvetti@isti.cnr.it

Keywords: Image mining, image algebras, medical image analysis, pattern recognition in image understanding, information technologies, automated diagnosis.

Abstract: The paper is devoted to the development and formal representation of the descriptive model of information technology for automating morphologic analysis of cytological specimens (lymphatic system tumors). The main contributions are detailed description of algebraic constructions used for creating of mathematical model of information technology and its specification in the form of algorithmic scheme based on Descriptive Image Algebras. It is specified the descriptive model of an image recognition task and the stage of an image reduction to a recognizable from. The theoretical base of the model is the Descriptive Approach to Image Analysis and its main mathematical tools. It is demonstrated practical application of algebraic tools of the Descriptive Approach to Image Analysis and presented an algorithmic scheme of a technology implementing the apparatus of Descriptive Image Algebras.

1 INTRODUCTION

The paper is devoted to the development and formal representation of the descriptive model of the information technology for automating morphologic analysis of cytological specimens of patients with lymphatic system tumors. The main contribution are detailed description of algebraic constructions used for creating of mathematical model of the information technology and its specification in the form of an algorithmic scheme based on Descriptive Image Algebras (DIA). We specify, in particular, the descriptive model of an image recognition task and the stage of an image reduction to a recognizable form.

The theoretical base of the model is the Descriptive Approach to Image Analysis (Gurevich, 2005) and its main mathematical tools –DIA, Descriptive Image Models (DIM) and Generating Descriptive Trees (GDT).

In a sense the results are continuation, specification and extension of the previous research. In (Gurevich, et al. 2007) we presented a brief introduction into the essential tools of the Descriptive Approach (DIA, DIM, GDT), the simplified model of an image recognition task based on multi-model image representation, a descriptive model of the information technology, and the descriptive and the structural schemes of the information technology. The state of the art and motivation were presented in our previous publications (Gurevich, et al. 2003, 2006, 2007).

Section 2 illustrates a simplified descriptive model of an image recognition task based on multi-model image representation. In section 3 we introduce operands and operations (and its operational (semantic) functions) of DIAs and

Descriptive Image Groups (DIG) necessary for constructing the algebraic model of the morphological analysis of lymphatic cell nucleuses. Section 4 presents a descriptive model of the information technology for automating morphologic analysis of cytological specimens of patients with lymphatic system tumors. The technology has been tested on the specimens from patients with aggressive lymphoid tumors and innocent tumor. The results are discussed in Section 4.

The main components of the technology are described via DIA tools and presented as an algorithmic scheme. The latter ensures a standard representation of technologies for intellectual decision making.

2 DESCRIPTIVE MODEL OF AN IMAGE RECOGNITION PROBLEM

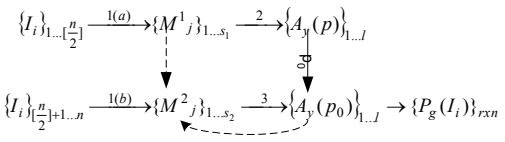
The Descriptive Approach provides the following model for an image recognition process (Gurevich, 2005):

$$\{I_i\}_{1...n} \to \{M_j\}_{1...s} \to \{A_y\}_{1...l} \to \{P_g(I_i)\}_{rxn} \quad (1)$$

$\{I_i\}_{1...n}$ - a set of initial images. $\{I_i\}_{1...n} \subset \bigcup_1^r K_g$, $\{K_g\}_{1...r}$ - a set of classes determined by an image recognition task, $\{M_j\}_{1...s}$ - a multimodel representation of each initial image $\{I_i\}_{1...n}$. An algorithm combination $\{A_y\}_{1...l}$ solves an image recognition problem, if it puts a set of predicates $\{P_g(I_i)\}_{rxn}$ into correspondence to the set of initial images, where predicate $P_g(I_i)=a_{ig}$ has the values: $a_{ig}=1$, if an image I_i belongs to a class K_g; $a_{ig}=0$, if an image I_i does not belong to a class K_g; $a_{ig}=\Delta$, if an algorithm combination does not establish membership of an image I_i to a class K_g.

Multi-model representation is generated by the set of GDT. Different ways for constructing multi-aspect image representations may use different types of GDT. An image representation becomes a multi-model one, if it is generated by different types of GDT.

This model including a training stage is as follows:



(2)

The descriptive models could be represented as algorithmic schemes containing 3 stages: 1) an image reduction to a recognizable form (an image model (models) construction); 2) training (adjusting parameters of chosen algorithms on a training set of images); 3) recognition (sequential application of chosen algorithms with adjusted parameters to each image under recognition). Construction of a multi-model representation is conceptually the same for both training set and recognition set; however, as it will be shown below, training and recognition process can ramify in stage 1. The latter consists of 2 sub-stages: 1(a) - construction of a multi-model representation for training set; 1(b) construction of a multi-model representation for recognition set. In accordance with chosen recognition algorithms the sub-stage 1(b) is executed together with sub-stage 1(a) (a case of the same multi-model representations for training and recognition sets), or it is executed after sub-stage 1(a) (the sub-stage 1(a) defines multi-model representations for recognition set), or it is executed after the stage 2. The latter is a case when recognition algorithm influences the choice of multi-model representations for a recognition set.

3 DESCRIPTIVE IMAGE ALGEBRAS

In this section we introduce operands and operations (and its operational functions) of DIAs and DIGs necessary for constructing the algebraic model of the morphological analysis of lymphatic cell nucleuses.

DIA 1 is a set of color images. ***The operands***: a set U of $\{I\}$ - a set of images $I=\{\{(r(x,y), g(x,y), b(x,y)), r(x,y), g(x,y), b(x,y) \in [0...M-1]\}, (x,y) \in X\}$, $M=256$ - the value of maximal intensity of a color component, n - a number of initial images, X - a set of pixels. ***The operations*** are algebraic operations of vector addition module M, vector multiplication module M and taking an integral positive part of multiplication module M by an element from the field of real numbers in each image point: *1)* $I_1+I_2=\{\{((r_1(x,y)+r_2(x,y)) \bmod M, (g_1(x,y)+g_2(x,y)) \bmod M, (b_1(x,y)+b_2(x,y)) \bmod M), r_1(x,y), r_2(x,y), g_1(x,y), g_2(x,y), b_1(x,y), b_2(x,y) \in [0...M-1]\}, (x,y) \in X\}$; *2)* $I_1 \cdot I_2=\{\{((r_1(x,y) \cdot r_2(x,y)) \bmod M,$

($g_1(x,y)\cdot g_2(x,y)$) mod M, ($b_1(x,y)\cdot b_2(x,y)$) mod M), $r_1(x,y)$, $r_2(x,y)$, $g_1(x,y)$, $g_2(x,y)$, $b_1(x,y)$, $b_2(x,y) \in [0...M-1]\}$, $(x,y) \in X\}$; 3) $\alpha I=\{\{([\alpha r(x,y)$ mod M], $[\alpha g(x,y)$ mod M], $[\alpha b(x,y)$ mod M]), $r(x,y)$, $g(x,y)$, $b(x,y) \in [0...M-1]$, $\alpha \in R\}$, $(x,y) \in X\}$. DIA 1 is applied to describe initial images and the multiplication operation of *DIA 1* is applied to describe segmentation of diagnostically important nucleus on images.

DIG 1 is a set of operations $sb((U,C)\rightarrow U')$ for obtaining a binary mask corresponding to an indicated lymphocyte cell nuclei, C - the information about the contours of indicated nucleus, a set U' - a subset of a set U. If an image point (x,y) belongs to indicated nuclei then $r(x,y)=g(x,y)=b(x,y)=1$, if a point (x,y) belongs to nuclei background, $r(x,y)=g(x,y)=b(x,y)=0$. ***The operands:*** Elements of DIG 1 are operations $sb((U,C)\rightarrow U') \in B$. ***The operations*** of addition and multiplication are introduced on the set of functions s*b* as sequential operations for obtaining a binary masks and their addition and multiplication correspondingly: 1) $sb_1(I,C)+sb_2(I,C)=B_1+B_2$; *2)* $sb_1(I,C)\cdot sb_2(I,C)=B_1\cdot B_2$. *DIG 1* is applied to describe a segmentation process.

DIG 2 is a set U' of binary masks. ***The operands:*** Elements of DIG2 are binary masks $B=\{\{(r(x,y), g(x,y), b(x,y)), r(x,y), g(x,y), b(x,y) \in \{0,1\}, r(x,y)=g(x,y)=b(x,y)]\}, (x,y) \in X\}$, $M=256\}$. ***The operations*** of addition and multiplication are operations of union and intersection correspondingly: 1) $B_1+B_2=\{\{(r_1(x,y) \vee r_2(x,y), g_1(x,y) \vee g_2(x,y), b_1(x,y) \vee b_2(x,y)), r_1(x,y), r_2(x,y), g_1(x,y), g_2(x,y), b_1(x,y), b_2(x,y) \in \{0,1\}\}, (x,y) \in X\}$; *2)* $B_1\cdot B_2=\{\{(r_1(x,y) \wedge r_2(x,y), g_1(x,y) \wedge g_2(x,y), b_1(x,y) \wedge b_2(x,y)), r_1(x,y), r_2(x,y), g_1(x,y), g_2(x,y), b_1(x,y), b_2(x,y) \in \{0,1\}\}, (x,y) \in X\}$. *DIG 2* is applied to describe binary masks.

DIA 2 is a set of gray scale images. ***The operands:*** A set V of $\{J\}$ – a set of images $J=\{\{gray(x,y)\}_{(x,y) \in X}, (x,y) \in [0,...,M-1]\}$. ***The operations*** are algebraic operations of *gray* functions addition module M, multiplication module M and taking an integral positive part of multiplication module M by an element from the field of real numbers in each image point: *1)* $J_1+J_2=\{\{(gray_1(x,y)+gray_2(x,y))$ mod M, $gray_1(x,y)$, $gray_2(x,y) \in [0..M-1]\}$, $(x,y) \in X\}$; *2)* $J_1\cdot J_2=\{\{(gray_1(x,y)\cdot gray_2(x,y))$ mod M, $gray_1(x,y)$, $gray_2(x,y) \in [0..M-1]\}$, $(x,y) \in X\}$; *3)* $\alpha J=\{\{[\alpha$ $gray(x,y)$ mod M], $gray(x,y) \in [0..M-1]$, $\alpha \in R\}$, $(x,y) \in X\}$. *DIA 2* is applied to describe separated nucleus on images.

DIA 3 – a set F of operations $f(U\rightarrow V)$ converting elements from a set of color images into elements of a set of gray scale images. ***The operands:*** elements of DIA 3 - operations $f(U\rightarrow V) \in F$; such transforms can be used for elimination luminance and color differences of images. ***The operations*** of addition, multiplication and multiplication by an element from the field of real numbers are introduced on the set of functions f as sequential operations of obtaining gray scale images and their addition, multiplication and multiplication by an element from the field of real numbers correspondingly: *1)* $f_1(I)+f_2(I)=J_1+J_2$; *2)* $f_1(I)\cdot f_2(I)=J_1\cdot J_2$; *3)* $\alpha f(I)=\alpha J$. *DIA 3* is applied to eliminate luminance and color differences of images.

DIA 4 - a set G of operations $g(V\rightarrow P_1)$ for calculation of a gray scale image features. ***The operands****: DIA 4* - a ring of functions $g(V\rightarrow P_1) \in G$, P_1 - a set of P-models (parametric models). ***The operations.*** Operations of addition, multiplication and multiplication by a field element are introduced on a set of functions g as operations of sequential calculation of corresponding P-models and its addition, multiplication and multiplication by a field element. *1)* $g_1(J)+g_2(J)=p_1(J)+p_2(J)$; *2)* $g_1(J)\cdot g_2(J)=p_1(J)\cdot p_2(J)$; *3)* $\alpha g(J)=\alpha p(J)$. *DIA 4* is applied to calculate feature values.

DIA 5 - a set P_1 of P-models. ***The operands:*** a set P_1 of P-models $p=(f_1, f_2,...,f_n)$, $f_1, f_2,...,f_n$ - gray scale image features, n - a number of features. ***The operations:*** 1) addition – an operation of unification of numerical image descriptions: $p_1+p_2=(f^1_1, f^1_2,...f^1_{n1})+ (f^2_1,f^2_2,...,f^2_{n2})= (f^3_1,f^3_2,...,f^3_{n3})$, n_3 – a number of features of P-model p_1 plus a number of features of P-model p_2 minus a number of coincident features of P-models p_1; p_2, $\{f^3_1,f^3_2,...,f^3_{n3}\} \subset \{f^1_1,f^1_2,...,f^1_{n1}, f^2_1,f^2_2,...,f^2_{n2}\}$ - different features and coincident gray scale image features of P-models p_1 and p_2; 2) multiplication of 2 P-models – an operation of obtaining a complement of numerical image descriptions: $p_*p_2=(f^1_1,f^1_2,...,f^1_{n1})*(f^2_1,f^2_2,...,f^2_{n2})=(f^4_1,f^4_2,...,f^4_{n4})$, n_4 - a number of significant features of unified P-model of models p_1 and p_2, $f^4_1,f^4_2,...,f^4_{n4}$ - significant features obtained after analysis of features of P-model p_1 and P-model p_2, $f^4_1, f^4_2,...,f^4_{n4}$ may not belong to $\{f^1_1, f^1_2,...,f^1_{n1}, f^2_1,f^2_2,...,f^2_{n2}\}$ and may consist from feature combinations; 3) multiplication by a field element - operation of multiplication of a number, a vector, or a matrix by an element of the field: $\alpha p=\alpha(f_1, f_2,...,f_n)=(\alpha f_1, \alpha f_2,..., \alpha f_n)$. *DIA 5* is applied to select informative features. The addition is applied for constructing joint parametric image representation. The multiplication is applied for reducing a set of image features to a set of

significant features. The multiplication by an element from the field of real numbers is applied for feature vector normalization.

DIA 6 - a set P_2 of P-models (P_2 includes feature vectors of the same length). ***The operands****:* a set P_2 of P-models $p(J)=(f_1(J),f_2(J),...,f_n(J))$, n – a number of features, $f_1(J),f_2(J),...,f_n(J)$ - gray scale image features, $f_1(J),f_2(J),...,f_n(J) \in R$. ***The operations*** of addition, multiplication and multiplication by a field element are introduced on the set P_2 as operations of a vector addition, multiplication and multiplication by a field element: (1) $p(J_1)+p(J_2)=(f_1(J_1),f_2(J_1),...,f_n(J_1))+(f_1(J_2),f_2(J_2),...,f_n(J_2))=(f_1(J_1)+f_1(J_2), f_2(J_1)+f_2(J_2),...,f_n(J_1)+,f_n(J_2));$ (2) $p(J_1)*p(J_2)=(f_1(J_1),f_2(J_1),...,f_n(J_1))*(f_1(J_2),f_2(J_2),...,f_n(J_2))=(f_1(J_1)\cdot f_1(J_2), f_2(J_1)* f_2(J_2),...,f_n(J_1)\cdot,f_n(J_2));$ 3) $\alpha p(J)=\alpha(f_1(J),f_2(J),...,f_n(J))=(\alpha f_1(J), \alpha f_2(J),...,\alpha f_n(J))$. *DIA 6* is applied to describe images reduced to a recognizable form.

Table 1 shows all DIA with one ring and DIG used for describing the algorithmic scheme for solving the task of cytological image recognition.

4 AN ALGORITHMIC SCHEME OF THE MORPHOLOGICAL ANALYSIS OF THE LYMPHOID CELL NUCLEUSES

The developed information technology will be described below and represented by the algorithmic scheme (2) which is interpreted by means of DIA, DIM and GDT.

4.1 Initial Data

A database (DB) of specimens of lymphatic tissue imprints (Fig. 1) was created to select and describe diagnostically important features of lymphocyte nuclei images. DB contains 1830 specimens of 43 patients, both specimen images and the contours of diagnostically important lymphocyte cell nucleus indicated by experts. The patients belongs to the following diagnostic groups: aggressive lymphoid tumors (de novo large and mixed cell lymphomas (CL), transformed chronic lymphatic leukemia (TCLL)), innocent tumor (indolent chronic lymphatic leukemia (CLL)).

Table 1: DIAs with one ring used for describing algorithmic scheme for solving the task of cytological image recognition.

	Ring elements	Ring operations	Purpose
DIA1	color images	algebraic operations of vector addition module *M*, vector multiplication module *M* and taking an integral positive part of multiplication module *M* by an element from the field of real numbers in each image point	description of initial images and segmentation process
DIG1	operations of obtaining the binary mask corresponds indicated lymphocyte cell nuclei	sequential operations for obtaining a binary masks and their addition and multiplication	description of segmentation process
DIG2	binary masks corresponds indicated lymphocyte cell nuclei	algebraic operations of union and intersection	description of binary masks
DIA2	gray scale images	algebraic operations of *gray* functions addition module M, multiplication module M and taking an integral positive part of multiplication module M by an element from the field of real numbers in each image point	description of separated nucleus on images
DIA3	operations reducing color images to gray scale images	sequential operations of obtaining gray scale images and their addition, multiplication and multiplication by an element from the field of real numbers	elimination luminance and color differences of images
DIA4	operations of image feature calculation	sequential calculation of corresponding P (parametric)-models and its addition, multiplication and multiplication by a field element	feature calculation
DIA5	P-models	image algebra operations (union, complement, multiplication by real number)	selection of informative features
DIA6	P-models	operations of a vector addition, multiplication and multiplication by a field element	image reduction to a recognizable form

Table 2: Database Statistics.

Diagnosis	Patient number	Image number	Nuclei number
CL	18	986	1639
TCLL	12	536	1025
CLL	13	308	2497
Total:	43	1830	5161

Footprints of lymphoid tissues were Romanovski-Giemsa stained and photographed with digital camera mounted on Leica DMRB microscope using PlanApo 100/1.3 objective (Fig. 1). The equivalent size of a pixel was 0,0036 mcm^2. 24-bit color images were stored in TIFF-format.

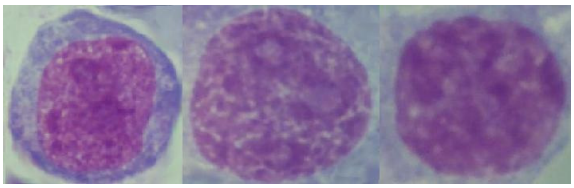


Figure 1: Specimen nucleus of patients with CL, TCLL and CLL diagnosis (from left to right).

4.2 Reducing an Image to a Recognizable Form

The initial images were divided into 2 groups: training image set $\{I_i\}_{1...\lceil \frac{n}{2} \rceil}$ and recognition image set $\{I_i\}_{[n/2]+1...n}$. The steps 1.1-1.6 of stage 1 "Reducing an image to a recognizable form") are described below as follows: description, step operands, step operations, results of step operation applying. It will be highlighted by letters 'a' and 'b' where processing of training and recognition sets differs.

Step 1.1: Obtaining Masks of Diagnostically Important Nucleus on Images. Application of segmentation algorithm is described by operands $sb((U,C)\rightarrow U')\in B$ of DIG1. An algorithm $sb((U,C)\rightarrow U')\in B$ is applied to initial images in order to obtain corresponding mask (equation 3).

$$\underset{DIA1}{\{I_i\}_{1...n}} \xrightarrow[1.1]{sb\in DIG1} \underset{DIG2}{\{B_j\}_{1...m}} \quad (3)$$

Step operands are initial images $\{I_i\}_{1...n}$ and contours of lymphocyte cell nucleus.

Step operation is an operation described by DIG1. Such description gives flexibility for using different kind of segmentation algorithms. The applied algorithm of threshold segmentation was supplemented by morphological processing of derivable nuclei images in order to obtain a corresponding mask.

Results of operation applying are binary masks $\{B_j\}_{1...m}$ represented as operands of DIG2.

Step 1.2: Segmentation of Diagnostically Important Nucleus on Images. The mask multiplication by an initial image gives indicated nuclei image (equation 4).

$$\underset{DIA1}{\{I_i\}_{1...n},\{B_j\}_{1...m}} \xrightarrow[1.2]{(\bullet)DIA1} \underset{DIA2}{\{M_1^T(I_{i(j)},B_j)\}_{1...m}} \equiv \{I_j^1\}_{1...m} \quad (4)$$

Step operands are initial images $\{I_i\}_{1...n}$ and binary masks represented as operands of DIG2.

Step operation is an operation of multiplication of 2 operands of DIA1. All initial images were multiplied by corresponding binary masks.

The results of the operation are T(transfomatonal)-models $\{I_j^1\}_{1...m}$ of initial images.

Step 1.3: Reducing Color Images to Gray Scale Images. To compensate different illumination conditions and different colors of stain the specimen images were processed before feature values calculation (equation 5).

$$\underset{DIA1}{\{I_j^1\}_{1...m}} \xrightarrow[1.3]{f\in DIA2} \underset{DIA3}{\{M_2^T(I_j^1)\}_{1...m}} \equiv \{I_j^2\}_{1...m} \quad (5)$$

Step operands are image models $\{I_j^1\}_{1...m}$.

Step operations are described by the elements of the DIA 2. Such representation gives flexibility for using different kinds of processing operations. Here the function f(U→V)∈F (DIA 2 element) has a form (I={{(r(x,y),g(x,y),b(x,y)),r(x,y),g(x,y),b(x,y) ∈[0..M-1]}$_{(x,y)\in X}$}): f(I)=J={{gray(x,y)}$_{(x,y)\in X}$, (x,y)∈[0...M-1]}, gray(x,y)=g(x,y)$\frac{2B}{M}$, B - an average brightness of a blue component of an initial RGB-image. The green tone in this case is the most informative.

The results of the operation are T-models $\{I_j^2\}_{1...m}$.

Step 1.4a: Feature Calculation on Constructed Image Models of the Training Set. To calculate different features the training set were processed by different operations of DIA 4 (equation 6) (m_1 equals to a number of segmented nucleus in training set).

$$\{\boldsymbol{I}_j^2\}_{1...m_1} \xrightarrow[1.4a]{\{g_1,g_2,...\}\in DIA4} \{\boldsymbol{M}_1^P(\boldsymbol{I}_j^2)\}_{1...m_1} \equiv \{\boldsymbol{M}_1^P(j)\}_{1...m_1} \quad (6)$$

DIA3 DIA5

Step operands are image models $\{\boldsymbol{I}_j^2\}_{1...m_1}$.

Step operations are described by the elements of DIA 4. Such representation gives flexibility for calculation of different features in order to obtain different P-models $\boldsymbol{M}_1^P(\boldsymbol{j})$ (elements of DIA 5). 47 features were selected for describing each of the images: the size of nucleus in pixels, 4 statistical features calculated on the histogram of nucleus intensity, 16 granulometric and 26 Fourier features of nucleus. $\boldsymbol{M}_1^P(\boldsymbol{j})$ is the vector with dimension 47 for each image model $\boldsymbol{I}_j^2$, j=1...m_1.

The results of the operation are P-models $\{\boldsymbol{M}_1^P(\boldsymbol{j})\}_{1...m_1}$.

Step 1.5a: Selection of Informative Features. This is an additional step of image model reduction. As it will be shown below the recognition algorithm was applied to both a full model $M_1^P(j)$ (j=m_1+1...m) and a reduced model $M_2^P(j)$ (j=m_1+1...m). At this step the constructed descriptions of images from the training set are studied for selecting the most informative features (equation 7).

$$\{\boldsymbol{M}_1^P(\boldsymbol{j})\}_{1...m_1} \xrightarrow[1.5a]{(+,\bullet,\alpha\bullet)DIA5} \{\boldsymbol{M}_2^P(\boldsymbol{M}_1^P(\boldsymbol{j}))\}_{1...m_1} \equiv \{\boldsymbol{M}_2^P(\boldsymbol{j})\}_{1...m_1} \quad (7)$$

DIA5 DIA6

The step operands are image models $\{\boldsymbol{M}_1^P(\boldsymbol{j})\}_{1...m_1}$.

Step operations are described by the elements of DIA 5. Operations of addition and multiplication are introduced for unificating and for reducing sets of image features to a set of significant features. Operation of multiplication by an element from the field of real numbers is introduced for normalization of feature vectors. Such representation gives flexibility for using different kinds of feature analysis to obtain a reduced set of features. Application of factor analysis to training image set detected 14 features with the largest loads in the first and second factor (Gurevich, 2006).

The results of the operation are P-models $\{\boldsymbol{M}_2^P(\boldsymbol{j})\}_{1...m_1}$ - a the vector with dimension 14 for each of image models $\boldsymbol{I}_j^2$, j=1...m_1.

Step 1.6b: Feature Calculation on Constructed Image Models of the Recognition Set. The steps 1.4 and 1.5 obtain a multi-model representation for training set. The step 1.6 is the step of feature values calculation for a recognition set (equation 8).

$$\{\boldsymbol{I}_j^2\}_{m_1+1...m} \xrightarrow[1.6b]{(g1,g2,...)\in DIA4} \{\boldsymbol{M}_1^P(\boldsymbol{I}_j^2)\vee \boldsymbol{M}_2^P(\boldsymbol{M}_1^P(\boldsymbol{I}_j^2))\}_{m_1+1...m} \equiv \{\Psi(\boldsymbol{j})\}_{m1+1...m} \quad (8)$$

DIA3 DIA6

Step operands are image models $\{\boldsymbol{I}_j^2\}_{m_1+1...m}$.

Step operations are described by the elements of DIA 4. To describe each image 47 or 14 features were selected.

The results of the operation are P-models $\{\Psi(\boldsymbol{j})\}_{m1+1...m}$ (note that the multi-model representation of images was constructed).

4.3 Training and Recognition

The class "Algorithms Based on Estimate Calculations" (AEC-class) were chosen as recognition algorithms since they can be conveniently represented by algebraic tools (Zhuravlev, 1998).

Initial Data. DIA 6 and its operands $\Psi(\boldsymbol{j}) \equiv \boldsymbol{M}_1^P(\boldsymbol{I}_j^2)\vee \boldsymbol{M}_2^P(\boldsymbol{M}_1^P(\boldsymbol{I}_j^2))$ (j=1...m) describe initial data for recognition algorithm *A* ($\Psi(\boldsymbol{j}) = (\psi_1, \psi_2, ..., \psi_n)$ - feature vector with a dimension n=47 or n=14, $\{\Psi(\boldsymbol{j})\}_{m1+1...m}$ - information about recognition set, $\{\Psi(\boldsymbol{j})\}_{1...m_1}$ - information about training set, $\{\boldsymbol{P}_g(\boldsymbol{I}'_j)\}_{rxm_1} = \{\boldsymbol{a}_{gj}\}_{rxm_1}$ - information about memberships of training set images to classes $\{\boldsymbol{K}_g\}_{1...r}$ ($\boldsymbol{a}_{gi} \in \{0,1\}$, r=3, $\{\boldsymbol{I}'_j\}_{1...m}$ - initial specimen images, one image for each indicated nucleus). Recognition algorithm

$A(\{\Psi(j)\}_{1...m_1}, \{a_{g_i}\}_{rxm_1}, \{\Psi(j)\}_{m1+1...m}) = \{a_{g_i}\}_{rx(m-m_1)} \in \{A_y\}_{1...l}$ solves an image recognition problem, $\{a_{gj}\}_{1...r}$ - an information vector of image model I_j^2 calculated by algorithm A (j=m_1+1…m).

The algorithms were applied to both full image models $M_1^P(j)$ (j=1…m, 47 features) and reduced image models $M_2^P(j)$ (j=1…m, 14 features).

Algorithmic Scheme. We described the main steps and elements of an algebraic model of information technology for automation of diagnostic analysis of cytological specimens of patient with lymphatic system tumors (Fig. 2):

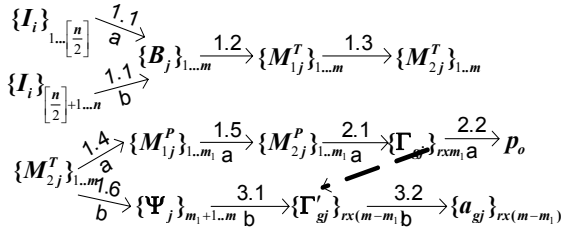


Figure 2: Algorithmic scheme of information technology.

Discussion of the Results. The elements of the technology were tested via software system «Recognition 1.0» (Zhuravlev, et al., 2005) including AEC-algoritms. It appeared that the best results are achieved by voting using all possible support sets, while automatic selection of support set cardinality and selection of support sets of fixed cardinality give lower precision.

Recognition rate for full feature set (Table 3) is 86,75%, while the rates differ for different recognition classes. High recognition rate for CLL (97,84%) is probably connected with innocent nature of CLL as opposed to CL (63,35%) and TCLL(84,51%) - the malignant cases. Thus, cells of CLL have evident distinctions from cells of other diagnoses, and cells of CL and TCLL are more similar to each other.

Table 3: The recognition rates for feature description consisted of 47 features.

Diagnosis	The number of correctly recognized cells	Total number of cells	The recognition rate
CL	693	820	84,51%
TCLL	325	513	63,35%
CLL	1221	1248	97,84%
Total cell set	2239	2581	86,75%

The recognition rate reduced feature set (14 features) decreased to 83,18% (Table 4). This feature set includes following features: size of nucleus in pixels, average by intensity histogram (statistic feature), numbers of elements with typical and minimal size of nuclei (granulometric features), 9 Fourier-features of nucleus.

Table 4: The recognition rates using reduced feature description consisted of 14 features.

Diagnosis	The number of correctly recognized cells	Total number of cells	The recognition rate
CL	626	820	76,34%
TCLL	300	513	58,48%
CLL	1221	1248	97,84%
Full cell set	2147	2581	83,18%

The software system «Recognition 1.0» (Zhuravlev, 2005), used for experimental investigation, includes effective realization of AEC methods and allows to apply them for practical task solution. It was experimentally verified that the best results are achieved by voting using all possible support sets, while automatic definition of support set capacity and definition of fixed support set capacity give lower precision.

5 CONCLUSIONS

The paper demonstrates practical application of algebraic tools of the Descriptive Approach to Image Analysis - it is shown how to construct a model of a technology for automation of diagnostic analysis on

images using. It is presented an algorithmic scheme of a technology implementing the apparatus of DIA. The paper solves a dual task: it presents technology by well structured mathematic model it shows how DIA can be used in image analysis application. The described techniques and tools will be used for creating software implementation of the technologies, its testing and performance evaluation.

ACKNOWLEDGEMENTS

This work was partially supported by the Russian Foundation for Basic Research Grants Nos. 05-01-00784, 06-01-81009, 07-07-13545, by Collaborative project “Image Analysis and Synthesis: Theoretical Foundations and Prototypical Applications in Medical Imaging” within agreement between CNR (Italy) and the RAS, by the project “Descriptive Algebras with one ring over image models” of the Program of Basic Research “Algebraic and Combinatorial Techniques of Mathematical Cybernetics” of the Department of Mathematical Sciences of the RAS, by the project no. 2.14 of the Program of the Presidium of the Russian Academy of Sciences “Fundamental Problems of Computer Science and Information Technologies”.

REFERENCES

Gurevich, I.B., 2005. The Descriptive Approach to Image Analysis. Current State and Prospects. In *Proceedings, 14th Scandinavian Conference on Image Analysis.* LNCS 3540, Springer-Verlag, Berlin Heidelberg.-P.214-223.

Gurevich, I., Harazishvili, D., Jernova, I., et al, 2003. Information Technology for the Morphological Analysis of the Lymphoid Cell Nuclei. In *Proceedings, The 13th Scandinavian Conference on Image Analysis*. LNCS 2749.-P.541-548.

Gurevich, I., Koryabkina, I., Yashina, V., Niemann, H., and Salvetti, O., 2007. An application of a Descriptive Image Algebra for Diagnostic Analysis of Cytological Specimens. An Algebraic Model and Experimental Study. In *Proceedings, The 2th International Conference on Computer Vision Theory and Applications, VISAPP 2007.* Volume Special Sessions, INSTICC Press.-P. 230-237

Gurevich, I., Harazishvili, D.V. Salvetti, O., et al, 2006. Elements of Information Technology for Cytological Specimen Image Analysis: Taxonomy and Factor Analysis. In *Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications*. Vol.16, No.1, Pleiades Publishing, Inc. 1. -P. 113-115

Zhuravlev, Yu.I., 1998. An Algebraic Approach to Recognition and Classification Problems. In *Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications, Vol.8.* MAIK "Nauka/Interperiodica".-*P.59-100*.

Zhuravlev, Yu.I., Ryazanov, V.V., and others, 2005. RECOGNITION: A Universal Software System for Recognition, Data Mining, and Forecasting. In *Pattern Recognition and Image Analysis*, *Vol. 15, No. 2*. Pleiades Publishing, Inc.-P.476-478.

REQUIREMENTS ENGINEERING TO AUDIT PRIVACY ISSUES IN MEDICAL AND HEALTH SOFTWARE

Miguel A. Martinez, Ambrosio Toval and Manuel Campos
Computer and Systems Departament, University of Murcia, Campus de Espinardo (Murcia), Spain
mmart@um.es, atoval@um.es, mcampos@dif.um.es

Keywords: Personal Data Protection, Audit, Requirements Engineering, Privacy, Reuse.

Abstract: In recent years, there has been a growing interest to guarantee that health organizations make a suitable treatment and protection of the personal data with which they deal in their daily activity. The privacy of personal data is regulated by law in many countries and is considered an important issue in a number of Quality Standards. This paper presents a systematic method to make an audit of the privacy in health sector software based on Requirements Engineering (RE). The aplication and validation of the method is ilustrated in a operative tool of report and clinical record management in the Intensive Care Unit (ICU) in a hospital.

1 INTRODUCTION

The Information Systems (IS) Audit is a discipline whose practice has increased considerably during the last few years. IS Audit is defined as the systematic process of gathering, grouping and evaluating evidences to determine whether an IS safeguards the assets, whether it maintains the integrity of the data, whether it effectively carries out the aims of the organization, and whether it uses the resources efficiently (Weber, 1988). A special type of audit within this discipline is the software audit. The purpose of auditing software is to verify that the software accomplishes the requirements (both functional and non-functional). A security audit is defined as "*an independent revision and examination of the registries and activities of a system in order to verify if the controls of the system are adapted to guarantee the fulfilment of the established policy and the operative procedures, to detect security problems, and to recommend possible changes in the control policy and procedures*" (ISO/IEC-7498-2:1989, 1989). A security audit can include many aspects, such as the protection level of the facilities or people, but this paper we will focus on the security related to data and information (privacy) of personal nature, that is a crucial aspect in the security of clinical IS and clinical software. More concretely, in this work we define a method to audit the aspects related to the privacy on clinical software, and we illustrate its use by means of an application to software for Electronic Clinical Record and report management for an Intensive Care Unit (ICU) that is being deployed in a hospital.

The method used to carry out the audit is based on SIREN, a general method of Requirements Engineering (RE) based on standards of this discipline (IEEE-Std.830-1998, 1998; IEEE-Std.1233-1998, 1998) and that is focused on the reusability of requirements. SIREN (*SImple REuse of software requiremeNts*) is a practical proposal to select and to specify the requirements of a software system. The key elements in SIREN are a spiral process model, requirements document templates and a reusable requirements repository which is organized by catalogs. The SIREN catalog related to privacy aspects is called PDP (Personal Data Protection) (Toval et al., 2002b).

At the moment, SIREN can be applied in four different ways:

1. As a method for RE (Toval et al., 2002a; Toval et al., 2002b) so that the fulfilment of the applicable norm (for example, on security and PDP) can be guaranteed from the beginning of the development of IS, by using suitable catalogs.
2. As guide and support to conduct an audit for determining the existence of security controls and its degree of fulfilment in an organization who deals with sensible data (Martinez et al., 2006).
3. As a method for auditing software (either developed by the organization or acquired) in operation.
4. As a method of consulting in the acquisition of

new software, so that it can be guaranteed that this software satisfies the expected level of security.

In this paper, we show a specialization of the method presented in (Martinez et al., 2006) that can be used in applications 3 and 4, that is, as a method for auditing and consulting. In order to conduct the audit in a simple and agile way, it is necessary to define a new method for auditing software that fulfils the standards of different disciplines, such as IEEE830 (IEEE-Std.830-1998, 1998) and CobiT (Control for Objectives Information Technologies) (CobiT, 2005), as well as the different laws in security and protection of personal data.

The rest of the paper is organized as follows: Section 2 describes briefly the method used for auditing. In Section 3, we describe the phase of practical application in our case study. In Section 4, some related works are presented and compared with our proposal. Section 5 describes the lessons learned after the application of the method proposed in the practical case. Finally, we enumerate the conclusions and the future work in this line.

2 A METHOD FOR AUDITING SOFTWARE BASED ON SIREN

The method proposed belongs to the scope of the Spanish legislation (LOPD, 1999; SMR, 1999), which is an adaptation of the European Union legislation (Directive-1995/46/CE, 1995). Nevertheless, the proposal is easily generalizable since this legislation has been adapted in a similar way by other European countries. For example, in Italy, the law that regulates the personal data protection is the *Italian Law 196*, of 30 of June of 2003, in Germany the counterpart is the *Federal Law of Data Protection* (BDSG), and in the United Kingdom, it is the *Data Protection Act* of 24 from October of 1998. In the United States, there is a sectorial conception of the law and has as source a mixture of legislation, regulation and self-regulation. Rights of information privacy in a variety of sectorial laws have been granted, like for example the law *The Privacy Act* of 1974, the *Fair Credit Reporting Act* of 1970, or the *Electronic Communications Privacy Act* of 1986.

In addition, our method has a direct correspondence with the referential frame of CobiT in its more recent version (CobiT, 2005), which is widely accepted by the international community of IS auditors. With this proposal, we intend to help to fulfil the CobiT Control Objectives that deal with privacy aspects, since the identification and verification of the fulfilment of the requirements related to these aspects is facilitated by the use of the requirements of the SIREN PDP catalog.

The method for auditing software based on SIREN contemplates the following phases (see Figure 1):

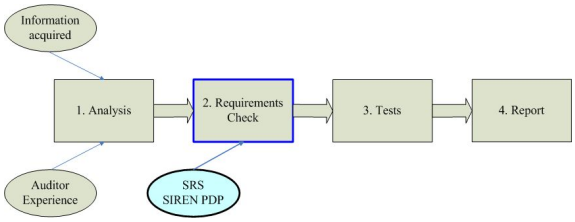


Figure 1: Phases of the method for auditing data protection in software.

Phase 1. - Analysis of the situation of the software. This phase consists of a first interview for establishing the scope of the audit so that an initial budget can be elaborated. The objective is to gather all kind of information about the treatment of the data used by the software. The information comes from two sources: on the one hand, the information facilitated by the development team of the software (manual of use, UML diagrams, etc.), and, on the other hand, from the experience acquired by the auditor after testing the software.

Phase 2. - Requirements verification with the SRS (Software Requirements Specifications) of the SIREN PDP catalog. The auditor verifies the fulfilment or breach of the requirements contained in the SRS of the catalog. Personnel responsible in the organization should support and facilitate, as much as possible, this verification. The verification consists of choosing those requirements of the SIREN PDP catalog that can be applied to the organization and of verifying whether they are fulfilled or not in the software. For example, if the software is going to be used in an organization to whom a high level of data protection is going to be demanded, according to the SMR (Security Measures Regulation), the auditor will extract from the SRS those requirements necessary to reach this protection level and will verify if these requirements are present in the tool. This extraction or filter of requirements of the catalog is possible thanks to the use of the meta-information associated to each requirement (in this case, through the attribute ”security level”).

Phase 3. - Execution tests. In this phase, the auditor must check the proper operation of the software once

it has been integrated in the system. In case that some of the evaluated measures are not fulfilled in the tool, the auditor must describe the risks that exist in the IS of the organization where the software is going to be used. In order to perform the tests, the *Software Test Specification* (STS) of the SIREN PDP catalog will be used. In this way, any person (even with little experience) could make the tests systematically.

Phase 4. - Preparation and writing of the final report. The product of the audit is a final report where all aspects from the evaluation are written. At least the report will have the following information:

- *Situation*: which describes briefly the resultant weakness after the analysis carried out in the software tool.
- *Threats*: where the possible risks which the software tool are exposed to, are enumerated.
- *Recommendations and action plans* proposed to the development team.

The aim of the SRS used for the audit is to gather the requirements about the functionality of the system, external interfaces, performance, design restrictions and attributes of software (portability, maintenance, security, availability and reliability). This specification of requirements has been made in agreement with the IEEE 830-98 Standard that is responsible for defining the characteristics and contents of a good software requirements specification.

A particular characteristic of the SIREN method is that in the requirements specification there exist what we call *parameterized requirements*. A parameterized requirement contains some parts that must be adapted to each application or system when they are reused in a concrete project. For example, the following requirement extracted from the SIREN PDP catalog "*the system will not conserve the data, once cancelled, so that the identification of the subject interested is not allowed during a period no inferior to [time in months] on the basis of which they had been successfully obtained or registered*", gives the analyst the possibility of choosing a period of suitable time to the necessities of the concrete project.

The SRS of the PDP catalog used for the audit presented in this work is currently composed by 48 requirements. In addition to the statement, each requirement of the catalog contains meta-information (attributes with information about each requirement) that enriches the requirement. Among the 18 attributes that have been defined up to now, we can point out the following ones: *source, exceptions, security level, motivation and fulfilment*.

3 CASE STUDY

In this work, we present a concrete application of the method to a software tool for clinical record management that deal with personal data and, therefore, needs a high protection level in its data. The tool used is called CH4 and has been developed by the Artificial Intelligence and Knowledge Engineering research group of the University of Murcia and is being deployed at the moment in the ICU of the *Hospital Universitary of Getafe (Madrid)*.

With the objective of taking advantage of the case study as an experience to validate the proposed method, we have decided to use a method of qualitative investigation in software engineering, denominated Action-Research (Baskerville, 1999) (the use of this method will be explained in Section 4.1).

CH4 has the aim of managing the clinical record as much as facilitating the daily work and communication between medical staff in an ICU. CH4 is centered in the process of patient management along their stay in the ICU by registering personal and clinical data, and allowing the staff to automatically generate reports. In this way, a documentary database with admittance, evolution and discharge report is created and available for its later access.

In the development of this report oriented tool, three scenarios have been considered: the admittance, the daily evolution and the discharge. In each one of them, data relative to tests (physical, complementary examinations, analytical explorations, classifications, ...) treatments and diagnoses can be introduced. The tool includes facilities, as the treatment profiles, to easy the management of the most habitual cases. Each diagnosis or problem can be related to the concrete results of tests, to the treatments or even to other problems, allowing in this way a traceability of all the aspects defined for a patient.

3.1 Design of the Experiment

The application of Action-Research gives rise to a cyclical process in which the different implied parts participate, examining the existing situation (which they consider of some problematic way) with the objective to change it and to improve it (Wadsworth, 1998). Action-Research is one of the few approaches valid to study the effects of specific alterations in development methodologies and maintenance of systems in human organizations (Baskerville and Wood-Harper, 1996). Following the terminology of Action-Research, in this case study the following participants are considered:

- The "researcher" is the research group in Software

Engineering of the University of Murcia.

- The "researched" object is the application of the PDP catalog in a method for auditing a software tool for clinical record management that needs a high protection level in the files of personal data that handles.
- The "critical reference group" (CRG), in other words, the one for which the research is carried out, is formed by the members of the development team of the tool, is the research group of Artificial Intelligence and Knowledge Engineering of the University of Murcia. According to Action-Research, the CRG must also participate in the research process, although less actively than the researcher.
- The "stakeholders" will be all those organizations who can benefit from the results of the research, in particular, the members themselves of the CRG and, in general, other groups whose development activities are similar to those of the tool audited.

In this research, a participative application of Action-Research has been made, in which the CRG puts into practice the recommendations made by the researcher, and shares its effects and results.

3.2 Audit of Software for Clinical Record Management in an ICU

Within IS Audit, we distinguished two basic types of audit: system audits and software audits. The first one corresponds to verifications on the own IS (an example of real application using method SIREN can be seen in (Martinez et al., 2006)). The second type is centered in verifying that the software tools accomplish to the requirements, both functional and non-functional (norms, laws, etc.). This work belongs to the second type, where the tool audited is a practical case of a software tool related to the clinical sector. We have used the SRS of the SIREN PDP catalog as guide to make the audit process shown in Figure 1.

After analyzing the information obtained in the different interviews with the development team of the tool made in Phase 1 of the audit, and after the test of the tool made by the auditor, we pass to Phase 2. Since we have at our disposal the SRS of SIREN PDP catalog, only two meetings were necessary to complete this phase. In these meetings, the fulfilment or breach of the requirements of SIREN PDP catalog has been checked one by one in the tool.

Once the two first phases of the audit method have been completed, the pertinent tests were made to verify the operation of the tool once it has been integrated in the IS.

Table 1: Results of the audit of the software.

	Software requirement
Fulfilled	22
Not fulfilled	10
Undetermined	16

Finally, all the results of the evaluation were written in a final report. This final report was given to the development team and they included in the tool the improvements and safety measures proposed in the audit. It is important to emphasize that implantation of such measures is not part of the audit, since a basic principle of the audits is that these finalize with conclusions and possible solutions, but never get to implement solutions as part of the audit.

In this way, after maintaining several interviews with the development team of the tool and the person in charge of security of the system, we have collected the data shown in Table 1.

According to the results of the Table 1, we see that the tool fulfils 45.8% of the referring requirements to the SRS of the SIREN PDP catalog. If we do not take into account the requirements that can not be applied to this tool (those marking like undetermined), a fulfilment of 68.7% with respect to the SRS of SIREN PDP catalog is obtained.

Some examples of requirements of the SIREN PDP catalog that were not fulfilled in the audited tool are the following ones:

- SRSL2. *The application will warn the user that a password needs to be changed.*
- SRSL6. *The [identification procedure] and [authentication procedure] will limit the possibility of repeatedly trying a non-authorized access to the application.*
- SRSL9. *The subsystem implementing the [identification procedure] and [authentication procedure], or other system related to this one, will log all accesses to the application. The log will consist of: user identification, timestamp of the access, file accessed, access type, and the result of the access.*
- SRSL14. *The application will allow the cancellation of the registered personal data (within the ten following days to its request by the interested part). The data will be accessible only by Public Administration and Courts for the investigation of possible responsibilities due to the treatment during the term of prescription of these responsibilities.*

The breach of these requirements is caused by diverse reasons. For example, requirement SRSL2 is

not fulfilled since the application did not provide a control to automatically warn the users about the necessity of changing the passwords. The breach of requirement SRSL6 put in serious danger the security of the application, because a user without permissions granted in the application could try to access many times; instead of only closing the session if an attempt of no authorized access takes place, the user login must be blocked and the attempts of no authorized access must be registered. Requirement SRSL9 is breached since the application did not consider registering the access information; as a solution, we proposed to use the database and application logging facilities to store the user accesses. Finally, requirement SRSL14 is not fulfilled since the application eliminates the data completely when a cancellation is required; instead of that, the data should be marked as unavailable or stored in a temporary database so that could be accessed in case they are required by the Public Administration.

4 RELATED WORK

In this section, we make a review of the most outstanding papers related to our proposal. In order to facilitate its reading, we have organized these works by topics, according to the essential related aspect treated. These topics are: 1) Personal data protection audits applied to the health sector; 2) Software tools audits; 3) Principles of the data protection in the European Union; 4) RE applied to the security field.

4.1 Personal Data Protection Audits Applied to the Health Sector

In (Sandhu and Samarati, 1996), an introduction to the personal data audit is provided, emphasizing its importance in the organizations that deal with them. Here, audit is understood like the process that gathers data on the system activities to analyze them in search of security violations. In (Hughes, 2005), the authors make a study about the relations between audits and research methods applied to the health sector of the United Kingdom. More concretely they apply Action-Research, which is a method with a special relevance in the health sector, where there is an important social factor. Nevertheless, in these papers, the approach is different from ours, since they do not audit a specific tool but the stored personal data in a system, and because neither application to a study case appears, nor the specific phases of an audit process are distinguished.

4.2 Software Tools Audits

Some of the software audit tools (in general, not only of the health sector) more used are GASP and WebCensus, both available in trial version in the BSA (*Business Software Alliance*) website[1], that is the main organization dedicated to the development of legal and safe computer science. These tools allow the auditors to identify and to track licensed and unlicensed software installed in a computer. Both tools are easy to use (WebCensus can be run through Web, without requiring installation by the user) and their high degree of accuracy has converted them in audit tools standard in numerous companies and governmental bureau worldwide. These tools have a different approach from the methodology that we propose, since they exclusively are limited to make a testing of the software installed in a system, contributing data related to the licenses and manufacturers, but they do not cover in any way the audit of software specific concerns (functional or non-functional requirements).

4.3 Principles of the Data Protection in the European Union

In (Van der Haak et al., 2003) and (Massacci et al., 2005), two practical applications about the use of personal data protection law in different European countries (concretely, Germany and Italy) are described. The first paper focuses on the identification of specific legal requirements related to data security and data protection of patients included in electronic clinical records. It is based on the set of German laws on data protection. The second paper presents a practical case of the application of a RE methodology for the fulfilment of the Italian legislation in privacy and data protection.

In these papers, authors do not provide the requirements engineers with a PDP requirements catalog (or similar) related to data privacy in a easily understandable language. So, the application of the data protection law in IS becomes a more tedious and difficult task for the requirements engineer.

The PDP requirements catalog that we have used (and improved) in this paper, is valid, with slight modifications, in any country of the European Union, since it is based on, among others, (Directive-1995/46/CE, 1995) and on (Directive-2002/58/CE, 2002). As (Lusignan et al., 2006) exposes, these directives are the base of the privacy laws of any European Union country. The authors show a table with the chronological order of the different treaties the

[1] www.bsa.org

fundamental principles of personal data protection of the European Union, and another table with a comparison of these principles (including general principles of the ethics in health computer systems).

4.4 Requirements Engineering applied to the Security Field

In the line of RE, we draw attention to the work by Firesmith (Firesmith, 2003), which provides examples and directives for requirements engineers to specify suitably security requirements. The different types of security requirements are identified and defined, among which privacy, security audit and physical protection requirements are highlighted. (Olvingson et al., 2002) presents a minimal data set for requirements elicitation in the area of public health. This minimal data set is a data collection that supports the elicitation of users' voices that later will constitute the foundation on which to identify the true requirements and describes the problems found by health professionals in their daily activities in countries like Sweden and the United States. Thus, it is possible to prioritize the requirements in an early phase of the construction of the system, e.g. the RE phase, and thus capture the most important characteristics to be implemented in IS. In these two papers, proposals are made considering thinking only RE and without following any RE methodology. On the contrary, we have followed the SIREN methodology and made a proposal for conduct audits.

In contrast to with the works described previously, our paper offers an integrated and repeatable systematic method to make an audit of software tools based on international standards of audit (CobiT) and good practices of Software Engineering (SIREN and international standards of RE). This method has been validated in a real study case where personal data are treated. In this way, our work complements other current proposals in the audit area but is new because allows the auditor to verify the fulfilment of privacy laws in software tools which deal with personal data.

5 LESSONS LEARNED

After carrying out this work we can identify three main lessons learned of our experience as software tools auditors.

On the one hand, this type of audits supposes a decisive aid to improve the used tools so that they adjust to the norms used as base; for example, in this case, the norms are the personal data protection laws. On the other hand, we have realized that thanks to the existence of a previous requirements catalog, we have been able to reduce the time dedicated to meetings and other activities of the audit. The interviews with the development team and board of directors of the organization (usually with very just a short time available) can be focused and guided directly to the crucial points that concern the audit.

Finally, we have detected weak points in the requirements catalog used in the audit. The inconsistencies consist fundamentally of the existence of ambiguities in the writing of certain requirements, which impede to the auditor team to make a firm decision on his fulfilment or breach in the software tool.

In relation to this last point, in addition to being able to detect what requirements were fulfilled or not in the current version of the developed tool, and thanks to the feedback obtained from the development team of the audit tool, ambiguous requirements and bad-written requirements in the SRS of the SIREN PDP catalog have been identified during the audit.

Of the 48 requirements that composed the SRS of the SIREN PDP catalog, 7 of them were identified by the development team of the audit tool, like candidates to a possible modification since they were not clear and precise enough. Some examples of these requirements, along with its corresponding improvement, are showed next:

SRSL2-Old. *The software will have to warn that a password needs to be changed.*

SRSL2-New. *In case that the authentication mechanism is based on the existence of passwords, the software will warn that a password needs to be changed, once finished the validity period set to [time in days].*

SRSL5-Old. *The software will allow that the Information Systems and data processing facilities can be subjected to an internal or external audit, at least every two years.*

SRSL5-New. *As much as every two years, the software will warn of the need to make an audit (internal or external) of the Information Systems and data processing facilities to verify the fulfilment of the Security Measures Regulations.*

SRSL26-Old. *The software will allow the people in charge of treatments of public and private ownership, as well as the organizations in which they group themselves, to formulate codes of ethics.*

SRSL26-New. In this case the requirement remains equal, it is necessary to add the following additional information in the attribute "*Rationale*" of the requirement: *The aim of codes of ethics is to establish the conditions of organization, operating conditions, applicable procedures, norms of security of the surroundings, use and obligations of the software and*

hardware implied in the treatment and use of the personal information, as well as the guarantees, in its scope, for the exercise of the people rights and its development norms.

6 CONCLUSIONS AND FURTHER WORK

Finally, we show the most important conclusions, obtained after carrying out this research:

- A systematic method has been extended to make a data protection audit to a software tool which deals with specially protected data. More specifically, we have applied the method to audit a tool for clinical record management in an ICU at a hospital that deals with sensible data.
- With the application of the proposed method, we helped to adapt the tool audited to the security measures and to the regulations demanded by law.
- An improvement of the PDP requirements catalog has been obtained (Toval et al., 2002b), which corresponds with one of the SIREN method phases. In this sense, the quality of the existing requirements has been improved: some requirements, identified as necessary or advisable, have been inserted in the PDP SIREN catalog.
- The use of "good practices" in Software Engineering considerably facilitates the subsequent audit work.
- As a result of the audit, we can provide a precise degree (in %) of fulfilment of a software tool with respect to the requirements document.
- With the application of the method, the treatment of legal requirements and the audit process implied is contemplated in the organization that uses the software tool. This process is considered important in Quality Standards like ISO 9001 or ISO 9004.
- The result of the audit provide the development team with a very useful and direct set of suggestions to be included in the software tool in order to accomplish the norms.

As limitation of our proposal, we can say that although our method is generalizable to other functional and non-functional software concerns, it is necessary to have a requirements catalog similar to the one used in this paper. At the moment, we have developed the following reusable requirements catalog:

- Personal Data Protection (PDP) (Toval et al., 2002b).
- Security in Information System (Toval et al., 2002a).
- Teleoperated Systems (Nicolas et al., 2006).

In lack of a predefined specific requirements catalog on the concern or concerns that we want to audit, some other alternative sources of broad acceptance could be used. For example, standard ISO/IEC 9126 (ISO/IEC-9126-1, 2001) could be used as guide for an audit of software quality.

As future work, we are already working on the development of a more specific requirements catalog of Electronic Clinical Record (ECR). This catalog will not only cover the legal aspects in the LOPD and the SMR but will also bring together the obligations imposed in the General Health Law (Law 14/1986) and the Law on the Autonomy of the Patient (Law 41/2002). Finally, we have developed a website where it is possible to consult the results obtained after audit, to download the SIREN PDP catalog and to request a change of the current catalog requirements.

ACKNOWLEDGEMENTS

This work has been partially financed by the Spanish Ministry of Science and Technology, project DEDALO (Desarrollo de sistEmas De calidad bAsado en modeLos y requisitOs), TIN2006-15175-C05-03, by the project FoMDAs (Formalizacion de (Meta-) Modelos y transformaciones en un marco MDA para el Desarrollo Automatico de SIW), URJC-CM-2006-CET-0387, by the project IDEATIO (Herramientas inteligentes para el control de la calidad asistencial en procesos de seguimiento integrado intra/extrahospitalario de pacientes: Una aproximacion basada en guias de practica clinicas), TIN2006-15460-C04-01, and by the Junta de Castilla-La Mancha (Spain), project DESERT (DEveloping Secure systEms through Requirements and Tools), PBC-05-012-3.

REFERENCES

Baskerville, R. L. (1999). Investigating information systems with action research. *Communications of the Association for Information Systems*, 2(3):4.

Baskerville, R. L. and Wood-Harper, A. T. (1996). A critical perspective on action research as a method for information systems research. *Journal of Information Technology*, 11(3):235–246.

CobiT (2005). *Control Objectives for Information and related Technology. 4th Edition*. http://www.isaca.org/cobit.htm.

Directive-1995/46/CE (1995). *Directive Num. 95/46/CE of the European Parliament and Council, dated October 24th: about People protection regarding the personal data management and the free circulation of these data. DOCE no. L281, 23/11/1995, P.0031-0050*.

Directive-2002/58/CE (2002). *Directive Num. 2002/58/CE of the European Parliament and Council, of July 12, 2002, relative to the processing of personal data and the protection of privacy in the electronic communications industry (Official Gazette of the European Union L 201 of 31.7.2002)*.

Firesmith, D. (2003). Engineering security requirements. *Journal of Object Technology (JOT)*, 2(1):53–68.

Hughes, R. (2005). Is audit research? the relationships between clinical audit and social research. *International Journal of Health Care Quality Assurance*, 18(4):289–299.

IEEE-Std.1233-1998 (1998). *Guide for Developing System Requirements Specifications. In Volume 1: Customer and Terminology Standards The Institute of Electrical and Electronics Engineers, Inc. IEEE Software Engineering Standards Collection*.

IEEE-Std.830-1998 (1998). *Guide to Software Requirements Specifications (ANSI). In Volume 4: Resource and Technique Standards The Institute of Electrical and Electronics Engineers, Inc. IEEE Software Engineering Standards Collection*.

ISO/IEC-7498-2:1989 (1989). *Information processing systems. Open Systems Interconnection. Basic Reference Model. Part 2: Security Architecture*.

ISO/IEC-9126-1 (2001). *Software Engineering - Product Quality - Part1: Quality Model.*

LOPD (1999). *Spanish Constitutional Law 15/1999, December 13th, on Personal Data Protection. BOE no. 298, 14/12/1999 (In Spanish)*. http://www.agpd.es.

Lusignan, S., Chan, T., Theadom, A., and Dhoul, N. (2006). The roles of policy and professionalism in the protection of processed clinical data: A literature review. *International Journal of Medical Informatics*.

Martinez, M. A., Lasheras, J., Toval, A., and Piattini, M. (2006). An Audit Method of Personal Data Based on Requirements Engineering. In *Proceedings of the 4th International Workshop on Security in Information Systems, (WOSIS'06), In conjunction with ICEIS'06, Paphos, Cyprus*, pages 217–231.

Massacci, F., Prest, M., and Zannone, N. (2005). Using a security requirements engineering methodology in practice: The compliance with the italian data protection legislation. *Computer Standards and Interfaces*, 27:445–455.

Nicolas, J., Lasheras, J., Toval, A., Ortiz, F. J., and Alvarez, B. (2006). A collaborative learning experience in modelling the requirements of teleoperated systems for ship hull maintenance. In *Proceedings of the Learning Software Organizations + Requirements Engineering (LSO+RE 2006), Hannover, Germany*, pages 71–80.

Olvingson, C., Hallberg, N., Timpka, T., and Greenes, R. (2002). Using the critical incident technique to define a minimal data set for requirements elicitation in public health. *International Journal of Medical Informatics*, 68:165–174.

Sandhu, R. and Samarati, P. (1996). Authentication, Access Control and Audit. *ACM Computing Surveys (CSUR)*, 28(1):241–243.

SMR (1999). *Spanish Royal Decree 994/1999, June 11th, by means of which the Security Measures Regulations of Automated Files which contain personal data is approved. BOE no. 151, 25/06/1999, page 24241 (In Spanish)*. http://www.agpd.es.

Toval, A., Nicolás, J., Moros, B., and Garcia, F. (2002a). Requirements Reuse for Improving Information Systems Security: A Practitioner's Approach. *RE J.*, 6(4):205–219.

Toval, A., Olmos, A., and Piattini, M. (2002b). Legal Requirements Reuse: A Critical Success Factor for Requirements Quality and Personal Data Protection. In *RE*, pages 95–103. IEEE Computer Society.

Van der Haak, M., Wolff, A., Brandner, R., Drings, P., Wannenmacher, M., and Wetter, T. (2003). Data security and protection in cross-institutional electronic patient records. *International Journal of Medical Informatics*, 70:117–130.

Wadsworth, Y. (1998). What is participatory action research? *Action Research International. Paper 2,*, 2004.

Weber, R. (1988). *EDP Auditing: Conceptual Foundations and Practice. 2nd Edition.* Mc Graw Hill.

SPECIFYING SECURITY POLICIES FOR ELECTRONIC HEALTH RECORDS

Felix Apitzsch
Fraunhofer Institute for Open Communication Systems FOKUS, Kaiserin-Augusta-Allee 31, Berlin, Germany
felix.apitzsch@fokus.fraunhofer.de

Stefan Liske, Thomas Scheffler, Bettina Schnor
Department of Computer Science, Potsdam University, Potsdam, Germany
{liske, scheffler, schnor}@cs.uni-potsdam.de

Keywords: Electronic Health Record (EHR), Security Policies, Digital Rights Language, Privilege Management and Access Control (PMAC).

Abstract: Sensitive data in electronic health records needs marking for special handling in order to maintain privacy. Person-centred records need mechanisms for individual and flexible marking. Policy mechanisms currently applied with shared health records in integrated care environments lack the ability to model complex privacy requirements. The paper examines two state-of-the-art policy languages for distributed processing environments such as web-services and digital rights management and describes how they can be applied with XML health records. Furthermore, it highlights the abstract concepts that need to be adopted and presents a distributed policy enforcement model.

1 INTRODUCTION

One of the major problems with integrated care depending on distributed care processes and distributed health IT is data interoperability between different IT systems. Centralised health care limited to one hospital and a single episode of illness can be handled by a single IT system or at least systems operated by a single authority. With integrated care, health data needs to be securely available at different locations in the care process.

The following paper describes how existing standards for the expression of data access policies for XML data can be applied to health IT-systems to protect Electronic Health Records (EHRs). We examine different EHR standards with respect to their support for access policies. Afterwards we analyse two policy languages that can be used to describe access control for distributed private medical data. [1]

Traditionally, security policies are first and foremost bound to a system, not to the resources. Even though there exist concepts to bind a security policy primarily to a resource (Wang, 2005), this approach has not been taken with EHRs yet. Existing access control methods are tailored towards a centralised data storage model. Consequently, medical records are stored in a medical database that is being accessed from authenticated users. Access control to the database items would be mediated by the database itself according to specific access control policies.

An important new aspect, that is specific to the domain of distributed health IT and the use of access control schemes, is the fact that there is no central policy management in place. Instead, data owners themselves should be able to set their own privacy preferences with the support of default settings and exemplary templates. These preferences then need to remain with the data as *sticky policies* (Mont et al., 2003)(Karjoth et al., 2003) throughout the ongoing data distribution process.

In a truly distributed system, medical data can migrate freely (e.g. on a patients smartcard (BMG, 2006)) and no centralised mechanism exists to protect access to the data. Therefore, the data has to be self-contained, which implies that it incorporates the necessary policies and protection mechanisms. A com-

[1] The security model proposed in this paper tries not to ignore or reinvent security or confidentiality concepts (Blobel et al., 2006) already described in ISO 22600, ISO 27799, ISO 21000, ISO 21731, EN 13606 and EN 13608, but to combine their views and to present some important common or extending concepts on an abstracting level.

mon security architecture has to be applied by any system involved in the care process and the relevant components have to comply with well-defined standards. We propose to mediate distributed data access with the help of explicit policy description languages.

The necessary authentication mechanisms as well as the mechanisms for the protection of the EHR (e.g. through encryption) are outside the scope of this paper, but the latter have been addressed in previous work (Apitzsch, 2007).

2 INTEGRATED CARE USE CASE

In this section we describe an integrated care use case that serves as the basis for the policy examples described in section 4. We present informal privacy requirements for the use case, as well as possible EHR XML-structures representing its data.

2.1 Shared Use of Medical Data

The following example of a fictional medical history is constructed to outline relevant characteristics for a study of privacy requirements. It shows how data of different sensitivity might be combined in one health record and different privacy requirements need to be be addressed by respective privacy policies.

```
Demographic Information:
  Name:            John Doe
  Date of Birth:   04.09.1977
  Place of Birth:  Berlin
Medical Entries:
01.06.1988 - (Dr. AB / XY Clinic)
  Diagnosis:  Cancer (Leukaemia)
  Treatment:  Planned chemo therapy
10.06.1988 - (Dr. AB / XY Clinic)
  Treatment:  Chemo therapy
  Result:     Everything OK
  Planned screening every year
01.07.1989 - (Dr. AB / XY Clinic)
  Treatment:  Cancer screening
  Result:     Everything OK
  [...]
30.08.2005 - (Dr. CD)
  Diagnosis:  Gonorrhoea infection
  Prescription: CIPROFLOXACIN 500 mg
15.04.2006 - (Dr. GH / UV Centre)
  Diagnosis:  Rheumatism
  Treatment:  Spinal X-ray
  Image:      <Attached X-ray>
  Prescription:  Novalgin
19.07.2006 - (Dr. IJ / ER)
  Treatment:  X-ray after traffic accident
  Image:      <Attached X-ray>
  Diagnosis:  Potential spinal injury
```

Listing 1: Exemplary Health Record.

Left out in the middle of the report are repetitive entries about the cancer aftercare. The rheumatism treatment includes a spinal X-ray that might be helpful for the medical opinion on a later potential spinal injury resulting from a car accident.
The health record addresses the patient's current situation with the latest entry for a potential spinal injury and shows useful X-rays from previous medical examinations. The further treatment process needs to add new data to the health record and needs to access parts of the old data.

Reflecting the idea of integrated care, the entries to the health record were made from different physician at different and independent institutions, all contributing to the ongoing treatment process for the patient. Accordingly, each involved computer system needs to access and process "foreign" data.

When Dr. IJ takes an X-ray to verify a potential spinal injury he might want to compare it to a prior spinal X-ray taken in the context of a rheumatism treatment. This raises the question how he knows about the prior data and how he can access it. A straightforward approach would imply that all medical documentation from every individual system is exported to a common EHR storage, allowing every party to access any data they need.

Since data from different systems is combined and leaves the individual security domains, this directly leads to the question whether everyone should have access to all medical data and how privacy of medical data can be preserved.

2.2 Confidentiality Requirements

The following parts of the paper will discuss technical details of a proposed security solution. Therefore, this section gives an informal description of exemplary security targets for the use case that need to be addressed in order to preserve privacy of shared medical data.

The listed confidentiality requirements are chosen to illustrate potential policy issues and even though commonly reasonable, do not necessarily offer maximum data protection.

Policy requirements might be different for every individual patient and under different situations. Consequently, data access privileges specified by the patient for the EHR are not static, but need to be dynamically extended or restricted.

In the example, the patient has no privacy concerns regarding the use of demographic data, rheumatism and the traffic accident, that can consequently be divulged to everyone. Data about cancer and the gonorrhoea infection are seen as sensitive and should not

have unlimited access. This requires the ability to effectively restrict data processing for everyone as chosen by the "owner" of the data, which usually is the patient.

In this use case example it is therefore not necessary to restrict the use of the rheumatism X-ray data. The sensitive data about the gonorrhoea infection, however, should only be visible to the EHR owner (patient). For the attending physician Dr. CD its visibility should be limited to the time period of treatment.

2.3 Data Representation in Electronic Health Records

Even though health records represented as free text are still in predominant use, standards for structured data representations in EHRs have been developed (e.g. (CEN/TS-15211, 2006)). Adding extra information on a meta-level, they not only ease data collection, data mining and reuse (cf. (Giere, 1986)) but additional allow the specification of security aspects at this level.

2.3.1 Genuine XML

Using XML to serialise the medical history described in Listing 2, the EHR would be structured by tags:

```
<healthRecord>
  <demographicInformation>
    <family_name>Doe</family_name>
    <given_name>John</given_name>
    <dayOfBirth>1977-09-04</dayOfBirth>
  </demographicInformation>
  <medicalHistory>
    <event date="1988-06-01" time="16:35:27">
      <diagnosis>Cancer</diagnosis>
      <treatment>Chemo therapy</treatment>
      <practitioner>DR. AB</practitioner>
    </event>
    [...]
    <event date="2005-08-30" time="10:35:27">
      <diagnosis>Gonorrhoea infection</diagnosis>
      <treatment>Ciprofloxacin 500 mg</treatment>
      <practitioner>DR. CD</practitioner>
    </event>
  </medicalHistory>
</healthRecord>
```

Listing 2: Health Record XML encoded.

The structure of the XML based EHR document needs to comply with an XML schema definition (Walmsley, 2004). This way, not only the structure of an EHR can be unified for simpler post processing, but specific tags or structures can be associated with semantic concepts.

Moving from free text towards structured data representation, the segment `<diagnosis>Cancer</diagnosis>` could be replaced with a specific tag `<cancerDiagnosis/>` or with another XML element of the type "cancerDiagnosis". This approach is used by HL7 CDA and EN13606 which represent the most commonly used and the latest development of EHR schemata, respectively.

2.3.2 Comments on HL7 CDA

HL7's Clinical Document Architecture (HL7, 2005) is not designed to support longitudinal records that cover complete accumulation of reports over time similar to the use case example above. A number of single documents could be used to represent the use case.

As mentioned for generic XML, CDA supports a standard XML schema defined structure as well as specific element tags associated with semantic concepts. CDA rel. 2 defines seven linked XML schema definitions, segmenting the XML structure into header and body, sections and entries. The semantic foundation for the contained elements is the HL7 Reference Information Model RIM (ISO/HL7-21731, 2006), complemented with fix vocabulary domains.

The information from the use case example can be represented by instances of the RIM classes Person, Procedure and Observation. A set of meta-data, e.g. a confidentialityCode, can be attached to each of these objects. Attached to an observation, the attribute specifies confidentiality rules limited to the object itself. When present with a person object, it refers to all entries related to the person. To allow for hierarchical confidentiality rules, a confidentialityCode may be specified at header, body, section, or entry level, each overwriting the more general.

Confidentiality rules are expressed using the vocabulary domain Confidentiality. It contains the following codes:

- low / normal / restricted / very restricted
- business / clinician / individual / substance abuse related / HIV related / psychiatry related / sexual and domestic violence related
- celebrity / sensitive / taboo

The vocabulary domain is a straightforward approach, but fairly incomplete, as it contains e.g. "HIV related", but no "cancer related" code. And although it defines the confidentiality intentions associated to the codes, they do not provide sufficient information to serve as machine processable security policies on their own. Nevertheless, they can be referred to as content types by higher level policies.

Additionally the RIM supports the use of roles, e.g. Patient, Employee, or special access roles. These might be referred to as types, similar to the confidentiality codes.

2.3.3 Comments on En 13606

Even though EN 13606 is a communication standard, it models the structure of the EHR with a two level approach, a reference model (EN-13606-1, 2007) and an archetype model (EN-13606-2, 2005). Using adequate archetypes, all data from the use case example can be represented and serialised in XML conforming to respective schemata (cf. openEHR (openEHR, 2007) which is an XML-based standard implementation close to EN 13606). In addition, using archetypes, which are bound to ontological concepts, semantic meta-information is available for all entries. This might aid the specification of security policies, as e.g. a cancer-related entry from the use case can be recognised as such, because it might be represented as a cancer-archetype.

On top of the reference and archetype model, EN 13606 defines its own security model (EN-13606-4, 2007). Using the concepts previously introduced in part 1 and part 2, part 4 defines an *access policy archetype*. (Medical) data that is represented as one or more *COMPOSITIONS* is complemented within a dedicated *Access policies FOLDER*. Each of these access policies itself is represented as a single *COMPOSITION*, whose archetype must conform to the specifications in part 4 of the standard. Additionally, a number of security categories are introduced with the security model:

- A Private entries shared with General Practitioner
- B Entries restricted to sexual health team
- C Entries accessible to administrative staff
- D Entries accessible to clinical support staff
- E Entries accessible to direct care teams
- F Private entries shared with several named parties
- G Entries restricted to prison health services

These categories can be used to mark data in an EN 13606 EHR at different levels of abstraction and thereby to assign respective policies, cf. Figure 1. Even though, this pragmatic approach gains some extra expressiveness through the use of flexible and extensible archetypes, it is still very limited in the kind of policies that can be specified. Although EN 13606 is still partly work in progress, it is to be doubted, that all confidentiality requirements of the use case described above, e.g. the delegation of rights, could be expressed without the integration of more comprehensive security standards. Therefore, a more general and flexible security model is described in the following section.

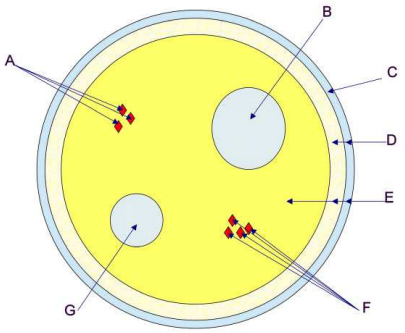


Figure 1: Access domains within 13606 EHRs(EN-13606-4, 2007).

3 EHR SECURITY MODEL

It is considered good practice to separate policies and mechanisms for access control and make the policy explicit. This allows independent implementation changes to enforcement mechanisms and opens the policy for external analysis and composition.

3.1 Policy (Description) Languages

From a mathematical point of view an access policy can be regarded as the set of all possible access decisions over the sets of subjects and objects supported by this policy. A definition of such an access policy is given by Woo (Woo and Lam, 1993): An authorisation policy is the 4-tuple (P^+, P^-, N^+, N^-) where each component is a subset of $\{(r,s,o) \mid r \in R, s \in S, o \in O\}$ over the set of subjects S, objects O and access rights R. P^+ and N^+ record the rights that are explicitly granted or denied. Whereas P^- and N^- record the rights that should not be explicitly granted or denied and are needed to define the semantics of policy composition.

Policy $A = (P^+, P^-, N^+, N^-)$ defines three authorisation relations for an authorisation request (r,s,o).

$$\begin{aligned} A\,grants\,(r,s,o) \quad &\text{iff} \quad (r,s,o) \in P^+ \\ A\,denies\,(r,s,o) \quad &\text{iff} \quad (r,s,o) \in N^+ \\ A\,fails\,(r,s,o) \quad &\text{iff} \quad (r,s,o) \notin P^+ \cup N^+ \end{aligned}$$

The policy representation as thus becomes irrelevant, as long as it can guarantee to be set-theoretical equivalent to the access matrix.

Policy description languages, such as the eXtensible Access Control Markup Language XACML

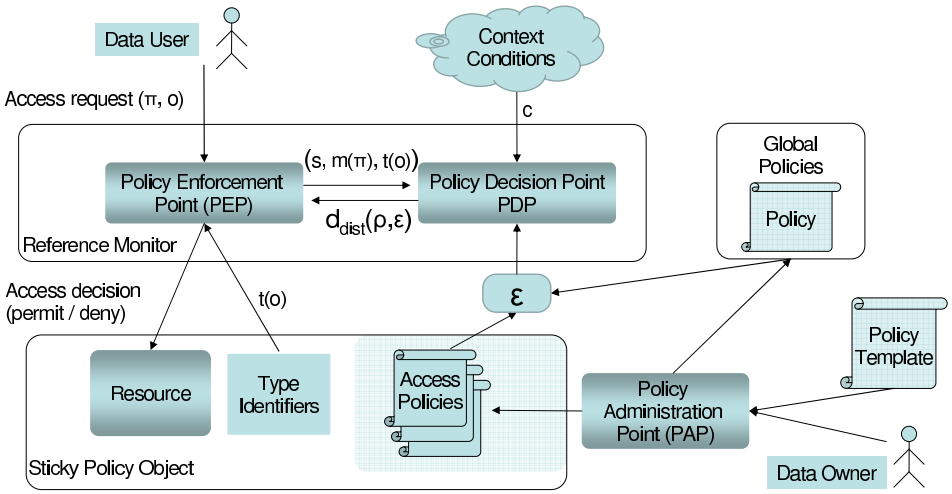


Figure 2: EHR Processing Model.

(XACML-2.0, 2005) or the eXtensible rights Markup Language XrML (ContentGuard, 2001) go one step further in their expressiveness and maintainability than simple *Access Control Lists* (ACL).

The ability of these languages to logically group objects and subjects, as well as the ability to evaluate environmental conditions, such as access-time, allow the creation of concise policies. These policies must be evaluated at the time of access in order to determine the access decision.

It is to be noted that with the use of these higher level policy languages the meaning of the term access decision is being extended. Since the rights that can be granted by a policy can reference complex processing concepts rather than simple access modes (read-/write), access control can be replaced by processing control. Consequently, these terms will be used synonymously, referring to the idea of processing control via a reference monitor.

3.2 Distributed Processing Control Architecture

A *Reference Monitor* is an instance that decides whether a process associated with the user $s \in S$ may execute a procedure $\pi \in \Pi$ on a resource $o \in O$ under the current system context $c \in C$. Therefore, it computes its decision for a request $\rho = (s, \pi, o, c) \in P = S \times \Pi \times O \times C$ as shown in equation (1).

$$d : P \rightarrow \{\text{grant, deny}\} \qquad (1)$$

In a distributed environment a reference monitor may not know S, Π, O, and C in advance. So, locally, d cannot be fully be defined. A decision d_{dist} needs to be based on an appropriate security (access) policy. To make this local decision for the same request ρ in a distributed environment, a rights expression ε needs to be evaluated together with the decision request. The expression ε may describe a number of policies that explicitly or implicitly address $S_o \times \Pi_o \times o \times C$ for a specific resource object o. A rights expression might use the concept of rights R instead of procedures Π. R represents generic action concepts, e.g. "view" or "amend", whereas Π contains specific program code blocks. Hence, a mapping $m : \Pi \rightarrow R$ needs to be defined for the actual execution context. Similarly, with $t : O \rightarrow T$ the rights expression may refer to types of objects $t(o)$ rather than the object o itself. Given a request $\rho \in P$ and a rights expression $\varepsilon \in E$, the reference monitor needs to compute

$$d_{dist} : P \times E \rightarrow \{\text{grant, deny}\} \qquad (2)$$

by matching the decision request against the represented policies.

In Figure 2 we show the complete processing model for distributed policy evaluation. The Data Owner creates a *Sticky Policy Object* at the *Policy Administration Point* (PAP), which he can also use to create global policies for all of his EHR data. The sticky data object combines EHR data with the actual Access Policy into a single XML file. It should be possible to support the policy creation process through the use of generic policy templates.

Data access from the Data User is only granted via the Reference Monitor which uses a modularised design, separating the *Policy Enforcement* from the *Policy Decision* component.

3.3 Sources of Authority

Only authoritative policies may be considered for the computation of d_{dist}. This means that not everyone is allowed to set permissions for a resource by defining and issuing arbitrary policies. Only the owner of a resource may author and authorise policies for it. Nevertheless, this authority could be extended by the owner to a third party, e.g. by making the extension part of a policy that grants the right to grant rights. The determination of authoritativeness of a policy is more complicated when privilege delegation or delegation of policy authoring is supported by the distributed processing control architecture.

Let $\varepsilon = (p_1 ... p_n)$ be a rights expression that consists of a set of policies by different authors $a(p_i)$ referring to a resource o. Further, let a_{owner} be the owner of o and let $v(\varepsilon)$ be a boolean function that is true, if and only if

$$\begin{array}{l} \forall p_i \in \varepsilon : a(p_i) = a_{owner} \vee \\ \exists p_{i-i} : p_{i-1} \text{ grants } a(p_i) \text{ to issue } (p_i) \end{array} \quad (3)$$

Then v is called verification of authority for ε. If $v(\varepsilon)$ is true, ε is called an authoritative rights expression and may be used to compute d_{dist}. With this definition the meaning of the term source of authority becomes clear, as all policy parts in an authoritative rights expression derive their authority from a single source a_{owner}. Any policy delegating rights or granting the right to issue new policies needs to be included in ε to allow the policy decision point to compute $v(\varepsilon)$.

4 XML POLICY LANGUAGES

This section compares two prominent policy description languages from the viewpoint of their applicability for the application domain.

4.1 XACML

The eXtensible Access Control Markup Language (XACML) is a declarative access control policy language and a processing model that is standardised by OASIS (XACML-2.0, 2005). The current version 2.0 was ratified in 2005.

4.1.1 Language Elements

PolicySet. The `<PolicySet>` element contains a set of `<Policy>` or other `<PolicySet>` elements and a *PolicyCombiningAlgorithm* to determine the joint evaluation of different elements.

Policy. The `<Policy>` element contains a set of rule elements and a *RuleCombiningAlgorithm* to determine the joint evaluation of the rules of the policy. It is the basis of an authorisation decision.

Combining Algorithms. XACML allows explicit positive and negative evaluation of rules (permit/-deny), as well as the combination of policies from different sources in a PolicySet for distributed policy generation. Combining algorithms are an essential part of the language specification. They are needed to derive an authorisation decision from potentially conflicting individual rules and policies. Standard combining algorithms are:

- Deny-overrides
- Permit-overrides
- First-applicable
- Only-one-applicable

Rule. The `<Rule>` element contains a policy expression that can be evaluated in isolation and provides the basic unit of policy management. The main components of a rule are its effect (permit/deny), target and potentially a condition that refines the applicability of the rule.

```
<Rule ...  Effect>
<Target>
...
</Target>
<Condition>
...
</Condition>
</Rule>
```

Target. The set of resources, subjects and actions to which rules and policies apply is called a target in XACML. Targets in policy elements define the scope of this element. If no restrictions have been made here the policy will have global scope.

Issuer/Delegation. As of version 2.0, XACML provides no mechanisms to describe a delegation policy as well as an issuer of a policy/delegation. In the current standard these have to be specified externally. Version 3.0 is currently in preparation and will add generic attribute categories and a policy delegation profile to the XACML specification.

4.1.2 Use Case Example

The XACML approach strictly separates authorisation policies and resources. Within XML-based re-

sources policies can be included and referenced via XPath (DeRose, 1999).

```
<HealthRecord>
 <Policy RuleCombiningAlgId="deny-overrides">
   <Rule Effect="Permit">
     <Target>
      <Subject><AnySubject/></Subject>
      <Resources>
        <ResourceMatch MatchId="xpath-node-equal">
        /HealthRecord/fileData </ResourceMatch>
      </Resources>
      <Action>view</Action>
     </Target>
   </Rule>
   <Rule Effect="Permit">
     <Target>
      <Subject>Dr. CD</Subject>
      <Resources>
        <ResourceMatch MatchId="xpath-node-equal">
    /HealthRecord/medHistory/event[Diagnosis="Gonorrhoea"]
        </ResourceMatch>
      </Resources>
      <Action>view</Action>
      <Condition FunctionId="date-less-than-or-equal">
        <Apply FunctionId="date-one-and-only"> </Apply>
        <AttributeValue>2002-03-22</AttributeValue>
      </Condition>
     </Target>
   </Rule>
 </Policy>
 <fileData>
    <demographicInformation>
   [...]
 </fileData>
 <medHistory>
   <event date="1988-06-01 16:35:27">
     <diagnosis>Cancer</diagnosis>
     [...]
   </event>
   <event date="2005-08-30 10:35:27">
      <diagnosis>Gonorrhoea infection</diagnosis>
      [...]
   </event>
 </medHistory>
</HealthRecord>
```

Listing 3: XACML Policy protected EHR.

4.2 XrML / MPEG-21

The origin of XrML is research on a "digital rights property language" (DPRL) by Stefik (Stefik, 1996). Version 2 of XrML developed at ContentGuard introduces a more generic approach to rights specifications. A revised edition was adopted by ISO as part 5 of the MPEG-21 standard. All subsequent statements refer to the MPEG-21 Version of XrML. Rights and other properties are represented by abstract concepts that are not bound to any context domain. Using XML namespaces, this basic XML structure can be extended with domain specific language elements replacing the abstract concepts. Therefore, it can be assumed that XRML can be adapted to the context of EHRs, a hypothesis that will be substantiated below.

4.2.1 Language Elements

License. A license in terms of XrML is a collection of grants allowing individuals to perform actions on specific resources. The `<license>` tag is the root of the XML tree and brackets all other relevant tags including `<grant>` or `<grantGroup>`, `<issuer>` and `<inventory>`. It does not contain any processible information itself. Alternatively, the non-obfuscated child nodes of a license can be replaced by an `<encryptedLicense>` which contains the same information, but needs to be decrypted for further processing.

Issuer. The `<issuer>` tag encloses a set of issuer-specific details about the circumstances under which he issues the license and a digital signature (Eastlake et al., 2002) for the license. With respect to the use case it could be the patient as a single source of authority signing the license.

Inventory. The `<inventory>` tag marks a part of a license that can be used to store anything referred to by a grant. By placing it in the inventory, redundancy, e.g. multiple principal or resource specifications, can be avoided when multiple grants refer to the same items. With respect to the use case, the inventory would be the place to include XML fragments of the EHR within `<digitalResource>` subsections, or to give an URI reference to an external EHR resource.

Grant or GrantGroup. Multiple subsections marked by `<grant>` tags may be present. A grant is the part of the license that specifies information relevant to decide whether a sub-procedure within a computer program should be executed with respect to the license (issuer's intention) or not. Like the `<license>`, the `<grant>` tag is only of syntactical nature. The relevant information is contained in the quartet of child nodes for principal, right, resource and condition, which all are conceptually abstract.

For each grant, additional pattern and delegation-control information can be stored in `<forAll>` and `<delegationControl>` tags respectively.

Therefore and with respect to the use case, the confidentiality requirements from Section 2.2 can be represented in the grant sections, including the intended delegation of privileges.

Principal, Right, Resource, Condition. Within each grant domain specific tags represent the abstract concepts principal, right, resource and condition. This means that there are no `<principal>` or `<right>` tags, but that these concepts can be substituted with domain specific tags, e.g. `<hpcHolder>` (for one specific holder of a health professional card) or `<compareImage>` (indicating the X-rays from the use case may be compared). These extensions to the abstract concepts are assembled in domain-specific XML-namespaces.

4.2.2 Use Case Example

The following code lists in an abridged form the required XrML language elements for the EHR use case example.

```
<license>
  <inventory>
    <digitalResource licensePartId="demoInfo">
      <XML>
        [Demographic Info]
      </XML>
    </digitalResource>
    [Cancer01]
    [...]
    [Gonorrhoea]
    [...]
    [Traffic Accident]
  </inventory>
  <grant>
    <export/>
    <digitalResource
     licensePartRef="demoInfo">
  </grant>
  [...]
  <grant>
    <[Dr. CD]/>
    <view/>
    <digitalResource
     licensePartRef="gonorrhoea">
    <notAfter>2006-02-30</notAfter>
  </grant>
  [...]
  <issuer> [Patient] </issuer>
</license>
```

Listing 4: XrML Policy protected EHR.

4.3 Comparison of XACML and XrML

Policy description languages differ in their ability to express certain concepts directly and efficiently as part of their language. Table 1 compares the support for different concepts in XACML and XrML.

Any XrML license is always granting. There are no denying syntax elements in the language. Therefore, the absence of a license never leads to false positive granting. This might be an advantage in distributed systems. XACML Policies instead have the ability to express negative authorisations and therefore can define explicit Policy/RuleCombiningAlgorithms for the inclusion of policies from different sources.

Table 1: Comparison between XACML and XrML.

	XACML	**XrML**
Explicit issuer	No element	Element
Condition support	Extensive	Extensive
Rights delegation	No	Yes
X500 naming	Supported	Not mandatory
X509 identities	Not supported	Supported
X509 attributes	Not supported	Not mandatory
Rule-signing	Indirectly through XML signatures	Directly supported
Encryption of content	Indirectly through XML encryption	Directly supported
Environment	Yes (time, etc.)	Yes (time, ticket, etc.)
Deny rules	Yes	No
Insertion of access rules	Easy [1]	Easy
Deletion of access rules	Easy [1]	Easy
Policy template support	Yes, through the use of `PolicySets`	No
Type identifiers	Not directly	Yes, (includes pattern matching based on XPath)

[1] might lead to changes in the policy (grant/deny rules)

Any requirement from Section 2.2 can be expressed with XrML and XCAML, because there is no limit to the integration of domain specific concepts. Domain independent requirements, e.g. the delegation of privileges, are featured by XrML itself.

Furthermore, the languages can express all elements of rights expression, as defined in the general security model in Section 3.2. The computation of $m(r)$ is in the responsibility of the policy enforcement point, referring to $t(o)$ is directly supported by XrML. e.g., it can refer to HL7 RIM attributes or EN13606 archetype using *resourcePatterns*.

5 CONCLUSIONS

In this paper we have examined the possibility to use existing XML policy languages that were developed

for digital rights management and the description of access policies for the protection of EHR data.

We foresee a need to mediate distributed data access, where data is stored, accessed and processed in a truly distributed fashion without the help of centralised policy mechanisms. Distributed data access, however, also requires a dedicated access control architecture, which we presented in Section 3 as a general model for access control in distributed processing environments, e.g. the medical IT environment described in the use case. Any concrete implementation of an policy enforcement mechanism can be analysed and compared with respect to this model.

The analysis of current EHR standards has shown that they are not ideally suited for reliable data protection and patient-controlled access restrictions. Instead, they should be used in combination with dedicated policy languages.

Section 4 presents two dedicated policy description languages that might be used to specify data access policies for EHR. A structural analysis and shortened example explains how these languages could be used. Even though a full policy description representing the use case could not be given for reasons of readability and length, their general applicability is shown. The two languages are compared face to face, outlining important differences when used for EHR protection.

An open issue and potential basis for further work is the formulation of a generic set of actions, rich enough for the fine-grained control over medical data in the workflow and simple enough for the patient to reliably apply in EHR policies.

REFERENCES

Apitzsch, F. (2007). Digital Rights Management for Electronic Health Records. In *Proceedings of CeHR International Conference 2007 (to appear)*.

Blobel, B., Nordberg, R., Davis, J., and Pharow, P. (2006). Modelling privilege management and access control. In *International Journal of Medical Informatics*, volume 75, pages 597–623.

BMG (2006). Die Spezifikation der elektronischen Gesundheitskarte. Bundesministerium für Gesundheit, Version 1.1.0, http://www.dimdi.de/static/de/ehealth/karte/index.htm.

CEN/TS-15211 (2006). Health informatics - Mapping of hierarchical message descriptions to XML. European Committee for Standardisation, http://www.cen.eu.

ContentGuard (2001). eXtensible rights Markup Language (XrML) 2.0, Specification.

DeRose, J. C. S. (1999). XML Path Language (XPath). W3C Recommendation, http://www.w3.org/TR/1999/REC-xpath-19991116.

Eastlake, D., Reagle, J., and Solo, D. (2002). RFC3235: Extensible Markup Language - XML-Signature Syntax and Processing. http://www.rfc-editor.org/rfc/rfc3275.txt.

EN-13606-1 (2007). Health informatics - Electronic health record communication - Part 1: Reference model. European Committee for Standardisation, http://www.cen.eu.

EN-13606-2 (2005). Health informatics - Electronic health record communication - Part 2: Archetypes. European Committee for Standardisation, http://www.cen.eu.

EN-13606-4 (2007). Health informatics - Electronic health record communication - Part 4: Security. European Committee for Standardisation, http://www.cen.eu.

Giere, W. (1986). *BAIK - Befunddokumentation und Arztbriefbeschreibung im Krankenhaus*.

HL7 (2005). HL7 Clinical Document Architecture, Release 2.0, Normative Edition.

ISO/HL7-21731 (2006). Health informatics - HL7 version Reference information model Release 1).

Karjoth, G., Schunter, M., and Waidner, M. (2003). Platform For Enterprise Privacy Practices: Privacy-enabled Management Of Customer Data. In *2nd Workshop on Privacy Enhancing Technologies (PET2002)*, volume Lecture Notes in Computer Science 2482, pages 69–84. Springer Verlag.

Mont, M. C., Pearson, S., and Bramhall, P. (2003). Towards Accountable Management of Identity and Privacy: Sticky Policies and Enforceable Tracing Services. In *Proceedings of the 14th International Workshop on Database and Expert Systems Applications*, page 377. IEEE Computer Society.

openEHR (2007). openEHR Release 1.0.1. http://www.openehr.org.

Stefik, M. (September 18th, 1996). The Digital Property Rights Language, Manual and Tutorial, Version 1.02. Technical report, Xerox Palo Alto Research Center, Palo Alto, CA.

Walmsley, D. C. F. P. (2004). XML Schema. W3C Recommendation, http://www.w3.org/TR/2004/REC-xmlschema-0-20041028/.

Wang, X. (2005). Desing Principles and Issues of Rights Expression Languages for Digital Rights Management. In *Proceedings SPIE, Conference on Visual Communications and Image Processing*, volume 5960, pages 1130–1141.

Woo, T. Y. C. and Lam, S. S. (1993). Authorizations in Distributed Systems: A New Approach. *Journal of Computer Security*, 2(2-3):107–136.

XACML-2.0 (2005). eXtensible Access Control Markup Language (XACML). OASIS-Standard, http://www.oasis-open.org/committees/xacml.

AUTOMATIC BREAST CONTOUR DETECTION IN DIGITAL PHOTOGRAPHS

Jaime S. Cardoso, Luis F. Teixeira
INESC Porto, Faculdade de Engenharia, Universidade do Porto, Portugal
jaime.cardoso@inescporto.pt, luis.f.teixeira@inescporto.pt

Maria J. Cardoso
Faculdade de Medicina, Universidade do Porto, Portugal
mjcard@med.up.pt

Keywords: Computer-aided medical system, breast cancer conservative treatment, breast contour, shortest path problem.

Abstract: Breast cancer conservative treatment (BCCT), due to its proven oncological safety, is considered, when feasible, the gold standard of breast cancer treatment. However, aesthetic results are heterogeneous and difficult to evaluate in a standardized way, due to the lack of reproducibility of the subjective methods usually applied. The objective assessment methods, considered in the past as being less capable of evaluating all aspects of BCCT, are nowadays being preferred to overcome the drawbacks of the subjective evaluation. A recent computer-aided medical system was developed to objectively and automatically evaluate the aesthetic result of BCCT. In this system, the detection of the breast contour on the digital photograph of the patient is a necessary step to extract the features subsequently used in the evaluation process. In this paper an algorithm based on the shortest path on a graph is proposed to detect automatically the breast contour. The proposed method extends an existing semi-automatic algorithm for the same purpose. A comprehensive comparison with manually-drawn contours reveals the strength of the proposed method.

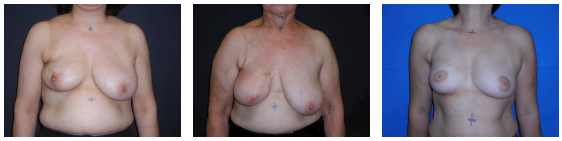
1 INTRODUCTION

Breast cancer conservative treatment (BCCT) has been increasingly used over the last twenty years due to identical survival results when compared to mastectomy, and a better cosmetic outcome. Although considerable research has been put into the oncological aspects of BCCT, diverse aesthetic results are common due to the difficult standardization of this type of treatment, stressing the importance of the aesthetic evaluation in institutions dedicated to breast cancer treatment so as to improve working practices.

The first and more generalized methods used for aesthetic evaluation of BCCT were the subjective appreciation of the patient, directly or through photographs, by one ore more observers (Harris et al., 1979). The categorization of the aesthetic result, subjectively estimated and combined by observers through this visual inspection relies on the complex interplay of multiple factors. Considering the subjectivity of any human decision, the obtained results are questionable and frequently the reproducibility values obtained among observers are only modest. In fact, this lack of reproducibility has been shown by others (Christie et al., 1996), which creates uncertainty when comparing results between studies. It has also been demonstrated that observers from different professional backgrounds evaluate cases differently, attaining even lower agreement results (Cardoso et al., 2005).

Objective methods of evaluation have emerged as a way to overcome these drawbacks of the subjective evaluation. Initially they consisted only in the comparison between the two breasts of simple measurements marked directly in patients or in photographs (Limbergen et al., 1989; Christie et al., 1996). The correlation of objective measurements with subjective overall evaluation has been reported by several authors (Christie et al., 1996). Trying to overcome the sense that objective asymmetry measurements were insufficient, other groups proposed the sum of the individual scores of subjective and objective individual indices (Al-Ghazal et al., 1999). More recently, a computer-aided medical system was developed to objectively and automatically perform the aesthetic evaluation of BCCT (Cardoso and Cardoso, 2007b). The

development of this system entailed the automatic extraction of several features from the photographs (Figure 1), capturing some of the factors considered to have impact on the overall cosmetic result: breast asymmetry, skin colour changes due to the radiotherapy treatment and surgical scar visibility. In a second phase, a support vector machine classifier was trained to predict the overall cosmetic result from the recorded features (Cardoso and Cardoso, 2007b).



(a) (b) (c)

Figure 1: Typical photographs.

In order to extract the identified relevant features from the image, the detection of the breast contour is necessary. In (Cardoso and Cardoso, 2007a) the authors describe a *semi-automatic* method for the detection of the breast contour. The user has to *manually* identify the two endpoints of the breast contour. Subsequently, the algorithm automatically finds the contour in-between. The algorithm has been implemented in a computer-aided medical system: the software automatically finds the contours, extracts relevant features and outputs a predicted overall cosmetic assessment (*excellent*, *good*, *fair*, or *poor*).

Here, we improve on the work of (Cardoso and Cardoso, 2007a) in two different directions. First, we present an algorithm for the automatic detection of the endpoints of the breast contour, thus eliminating any user input from the process. Therefore a totally automatic breast contour detection is achieved. Next, we provide a thorough evaluation of the performance of the proposed method against manually-drawn breast contours. Standard metrics are employed to compare two contours.

Before presenting the proposed approach, and for completeness, we recover the framework for breast contour detection between two *known endpoints* of (Cardoso and Cardoso, 2007a). Then, in section 3 we detail how to automatically find the endpoints of a breast contour. Examples are provided and a performance analysis is conducted in section 4. Finally, in section 5, we conclude the paper and present possible directions of future work.

2 A SHORTEST PATH APPROACH TO CONTOUR DETECTION

When knowing the two endpoints of the breast contour, we are left with the problem of finding the path between both endpoints that goes through the breast contour. As the interior of the breast itself is essentially free of edges, the path we are looking for is the shortest path between the two endpoints, **if** paths (almost) entirely through edge pixels are favoured. More formally, let s and t be two pixels of the image and $\mathcal{P}_{s,t}$ a path over the image connecting them. We are interested in finding the path $\mathcal{P}$ that optimizes some predefined distance $d(s,t)$. This criterion should embed the need to favour edge pixels.

In the work to be detailed, the image grid is considered as a graph with pixels as nodes and edges connecting neighbouring pixels. Therefore, some graph concepts are in order.

2.1 Definitions and Notation

A *graph* $G = (V,A)$ is composed of two sets V and A. V is the set of nodes, and A the set of arcs (p,q), $p,q \in V$. The graph is *weighted* if a weight $w(p,q)$ is associated to each arc, and it is called a *digraph* if the arcs are directed, i.e., $(p,q) \neq (q,p)$. A path from p_1 to p_n is a list of unique nodes $p_1, p_2, \ldots, p_n$, $(p_i, p_{i+1}) \in A$. The *path cost* is the sum of each arc weight in the path.

In graph theory, the shortest-path problem seeks the shortest path connecting two nodes; efficient algorithms are available to solve this problem, such as the well-known Dijkstra algorithm (Dijkstra, 1959).

2.2 Proposed Algorithm

If the weight assigned to an edge captures the intensity of the contour of the adjacent pixels, finding the best contour translates into computing the minimum accumulated weight along all possible curves connecting s and t:

$$d(s,t) = \min_{\mathcal{P}_{s,t}} \sum w(p,q). \quad (1)$$

Note that, if we ignore the weight component, we are simply computing the regular Euclidian distance between s and t along the path $\mathcal{P}_{s,t}$ (which will be a straight line for the shortest path).

Therefore, to detect the breast contour in (Cardoso and Cardoso, 2007a) it is proposed a two step approach:

1. Apply an edge detector to the original image (other operations, replacing the edge detector, will

be proposed latter in this work). The resulting binary image enhances the points of interest.

2. Detect the breast contour on the edge image, by finding the shortest path between the two endpoints.

We now detail this second step.

Starting by modelling the edge image as a graph, correspond a node to each pixel. Connect two nodes with an arc on the graph iff the corresponding pixels are neighbours (8-connected neighbourhoods) on the image. The weight of each arc is a function of pixels values and pixels relative positions — see Figure 2:

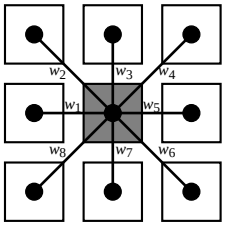


Figure 2: Arc weight between two pixels.

$$w_i = \begin{cases} f(p,q_i) \text{ if } q_i \in \text{4-connected neighbourhood of } p \\ h(p,q_i) \text{ if } q_i \notin \text{4-connected neighbourhood of } p \end{cases}$$

In this work we set $h(.,.) = \sqrt{(2)}f(.,.)$.

Now, to favour paths through edge pixels, one can set

$$f(p,q) = \begin{cases} c_1 \text{ if both p and q are edge pixels} \\ c_2 \text{ otherwise} \end{cases},$$

with $c_2 > c_1$. In this work c_1 and c_2 were experimentally determined as 2 and 32, respectively. Note that c_1 must be set greater than zero, to also favour the smallest path, when more than one exists through edge pixels only. Finally, the solution to the shortest path problem will yield the intended breast contour.

2.3 Algorithm Generalization

The proposed paradigm can be conveniently generalized. The application of an edge detector in the first step can miss to detect segments of the breast contour. This is especially true for women with small breasts (leading to weak contours) or when the breast is severely deformed with the excision of a large sample of tissue. A natural improvement is to replace the binary image outputted by the edge detector with a richer gradient image. Now, the shortest path algorithm should try to follow pixels with high gradient values. Thus, the $f(.,.)$ and $h(.,.)$ functions have to be properly generalized. A simple strategy is to set $f(p,q) = \hat{f}\big(255 - \min(grad(p), grad(q))\big)$, where $\hat{f}(.)$ is a monotonically increasing function. Note that this more general setting has the binary case as a particular instantiation. To summarize, the proposed general framework to find the contour between two endpoints encompasses:

- A gradient computation of the original image. In a broader view, this can be replaced by any feature extraction process that emphasizes the pixels we are seeking for.
- Consider the gradient image as a weighted graph with pixels as nodes and edges connecting neighbouring pixels. Assign to an edge the weight $w_i = f(p,q)$ or $w_i = h(p,q)$, as described before.

The gradient model adopted in the experiments reported latter is based on the Sobel operator. The Sobel operator is applied on the x and y directions; from the computed values, S_x and S_y, the magnitude of the gradient is estimated as $z = \sqrt{S_x^2 + S_y^2}$. Costs were assigned based on an exponential law:

$$\hat{f}(z) = \alpha \exp(\beta z) + \delta, \quad \alpha, \beta, \delta \in \Re$$

The parameters α, β, δ were chosen to yield $\hat{f}(0) = 2$ and $\hat{f}(255) = 32$, leading to the same range of costs as the binary model. The third degree of freedom was experimentally tuned. The adopted transformation was (see also Figure 3) $\hat{f}(z) = 0.15\exp(0.0208z) + 1.85$.

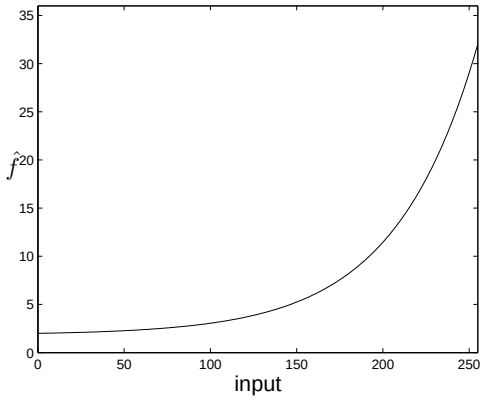


Figure 3: Transformation $\hat{f}(.)$.

3 AUTOMATIC DETECTION OF ENDPOINTS

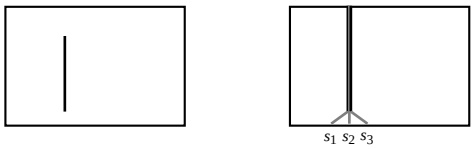
The challenge now is to automatically extract the endpoints. In order to address this key problem we will

assume, very reasonably, that the photo contains only the torso of the patient.

The position of the *external* endpoint of the breast contour can be assumed at the point of the body where the arm contour intersects the trunk contour. However, because patients are in the arms-down position, the arm's contour is almost indistinguishable from the trunk's contour. Therefore, we define the external endpoint of the breast contour as the highest point of the trunk contour.

The approach just delineated requires first the detection of the trunk contour. That may be searched among the strongest lines of gradient with approximate vertical direction.

In order to get a first intuition of the general result, it is instructive to explore a basic example. Consider in Figure 4(a) the gradient intensity, with black pixels representing high intensity values.



(a) A single vertical line of nonzero gradient.

(b) The shortest path between some points on the bottom and the top margin.

Figure 4: A first exemplificative example.

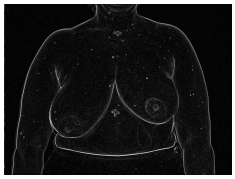
In Figure 4(b) the shortest paths between starting points s_i on the bottom row and the *whole* top row are traced. The distance between a *fixed* point on the bottom row and the whole top row can be generally formulated as the distance between a point s and a region Ω. The distance from a pixel s and a region Ω is given by

$$d(s,\Omega) = \min_{t\in\Omega} d(s,t), \qquad (2)$$

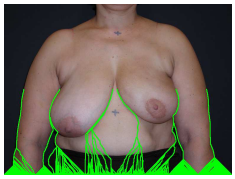
where $d(s,t)$ was defined previously. All the traced paths got attracted by the vertical line.

Phase 1: we propose to apply this procedure to the bottom half of our photographs:

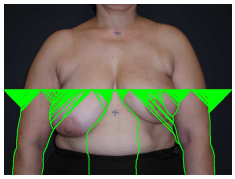
1. Compute the gradient of the image (see Figure 5(a)).
2. Compute the shortest path between each point in the bottom row and the *whole* middle row of the gradient image (see Figure 5(b)).
3. Compute the shortest path between each point in the middle row and the *whole* bottom row of the gradient image (see Figure 5(c)).
4. Discard all paths except for those common to steps 2 and 3 (see Figure 5(d)).
5. Discard paths with a cost superior to half of the maximum possible cost (see Figure 5(e)). Finally, the trunk contour is defined as the two contours closest to the middle of the photograph.



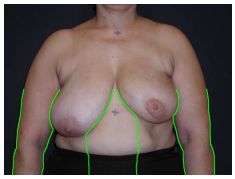
(a) Gradient.



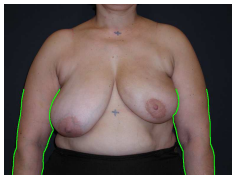
(b) Shortest paths from a pixel s in the bottom line and the whole middle row Ω_1, superimposed on the original image.



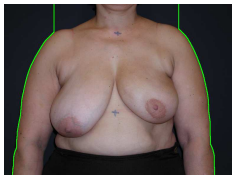
(c) Shortest paths from a pixel s in the middle row and the whole bottom row Ω_2, superimposed on the original image.



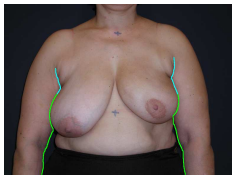
(d) Strong paths between the middle and bottom rows.



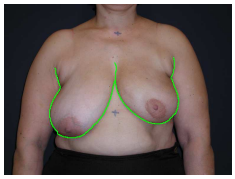
(e) Selected paths, superimposed on the original image.



(f) Strong paths between the top and bottom rows.



(g) The external endpoint of the breast contour is the highest point of the shortest path.



(h) Totally automated breast contour.

Figure 5: Results for a real photograph.

It is appropriate to introduce now the concept of *strong path*.

Definition. A path $\mathcal{P}_{s,t}$ is a strong path between regions Ω_1 and Ω_2 if $\mathcal{P}_{s,t}$ is the shortest path between $s \in \Omega_1$ and the whole region Ω_2, and $\mathcal{P}_{s,t}$ is the shortest path between $t \in \Omega_2$ and the whole region Ω_1.

With this definition, steps 2, 3 and 4 are just the computation of the strong paths between the middle and bottom rows.

At the end of this phase we have already the position of the two trunk contours, but we have stopped the process at the middle of the image. Perhaps it is important to stress that if the process was conducted between the bottom and top rows, the trunk contours would be lost, as the only strong paths between the top and bottom rows would be the external silhouette of the patient (see Figure 5(f)) .

Phase 2: to determine the top of the trunk contour, we need to continue the path produced in phase 1 until a certain condition is met. Towards that end, we propose to find the shortest path between the ending point of the strong contour found in phase 1 and row $R_i, R_i = middle\_row, \cdots, top\_row$. We select the highest row for which the shortest path does not contain a long sequence (`LENGTHTHRESHOLD`) of consecutive pixels with low gradient (`GRADIENTTHRESHOLD`). Figure 5(g) illustrates the results obtained for the exemplificative photograph (`LENGTHTHRESHOLD` and `GRADIENTTHRESHOLD` were set to 12 and 48, respectively).

Before applying the algorithm presented in Section 2 to compute the breast contour we need also the internal endpoint of the breast contour. This was estimated simply as the middle point between the two external endpoints. Finally, the computation of the breast contour yields the result presented in Figure 5(h).

4 RESULTS

The methodology proposed in this paper was assessed on a set of photographs from 120 patients. The photographs were collected in three different institutions in Portugal. All patients were treated with conservative breast surgery, with or without auxiliary surgery, and whole breast radiotherapy, with treatment completed at least one year before the onset of the study. Breast images were acquired employing a 4M pixel digital camera. A mark was made on the skin at the suprasternal notch and at the midline 25 cm below the first mark (see Figure 1). These two marks create a correspondence between pixels measured on the digital photograph and the length in centimetres on the patient.

In order to investigate the possibility of defining an automated method of detecting the breast contour, a set of patients with known breast contour was required. Since, ideally, the automated method should correlate coherently with human assessment, eight different observers were asked to manually draw the contours. A software tool was developed specifically to assist on this job. The user defines the contour by positioning seventeen control points of cubic splines, see Figure 6.

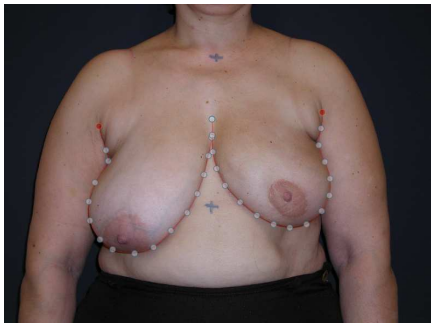


Figure 6: Software for manual breast contour definition.

Before applying the proposed algorithm, each image was downsized to a constant width of 768 pixels, while keeping the aspect ratio. This improves the computational performance of the implementation of the software, without degrading the quality of the final result.

Figure 7 shows the evolution of the error when estimating the external endpoints' position of the breast contour. The error in pixels was scaled to centimetres with the help of the marks made on the skin of the patient. Table 1 summarizes the results.

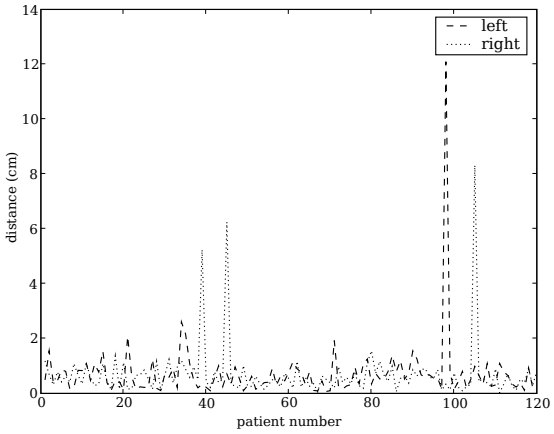
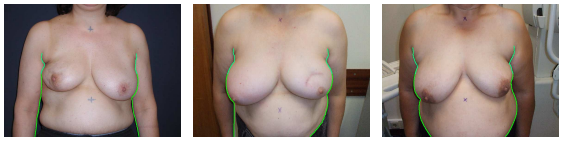


Figure 7: Evolution of error (cm) in the position of the external endpoints of the breast contour over 120 photographs.

Table 1: Mean, standard deviation and maximum value of errors in the position of the endpoints.

Error (cm)	mean	std dev	max
left endpoint	0.7	1.1	12.1
right endpoint	0.7	1.0	8.3
total	0.7	1.1	12.1

It can be observed that the proposed algorithm has a very interesting performance. The average error is quite low, less than 1 centimetre. Figure 8 shows some of the photographs for each the algorithm worked satisfactorily. It represents the result after the two phases of the algorithm. The highest point of the trunk contour provides the detected external endpoint of the breast contour. It is visible in patient #35 that the algorithm is robust against cluttered background. It is also visible in Figure 8 that, although the strong paths detected in phase 1 do not always correspond exactly to the trunk contour, the algorithm is still able to successfully detect the endpoints.



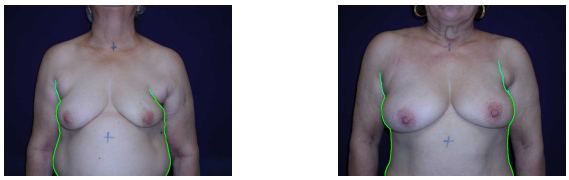
(a) Patient #05. (b) Patient #22. (c) Patient #35.

Figure 8: Selected successful results.

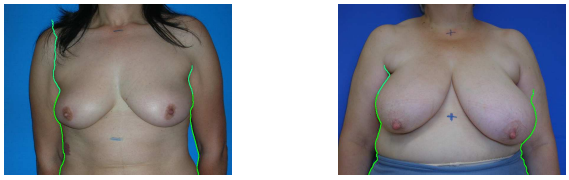
Nevertheless, four endpoints were clearly misplaced. These results, displayed in Figure 9, bring to light some of the limitations of the current state of the proposed approach. In patients #39, #45 and #105 the shortest path followed a 'wrong' contour, misplacing the endpoint. With patient #98 the long hair created a false 'path' till the top of the image.

After the automatic detection of the endpoints, the algorithm detailed in Section 2 to find the breast contour was applied to the photographs. Different scenarios can take place:

1. The endpoints were successfully located and the breast contour is correctly found. This desirable result is illustrated in Figure 10(a).
2. The endpoints were successfully located but the algorithm misses to follow adequately the breast contour (see Figure 10(b)).
3. The endpoints were poorly located but the algorithm rapidly finds and tracks the right breast contour (see Figure 10(c)).
4. The endpoints were poorly located and the breast contour is incorrectly tracked (this scenario did not occur in the experimental set of photographs).

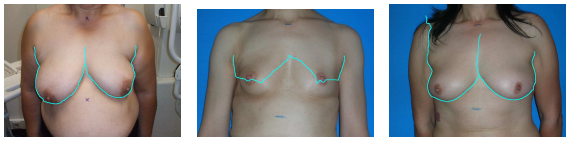


(a) Patient #39. (b) Patient #45.



(c) Patient #98. (d) Patient #105.

Figure 9: All poor results.



(a) Patient #35. (b) Patient #114. (c) Patient #98.

Figure 10: Breast contour results.

In a last set of experiments, the quality of the breast contour tracking algorithm was assessed. Instead of a simple subjective evaluation as provided in (Cardoso and Cardoso, 2007a), we conducted a complete objective evaluation, based on the hausdorff and the average distances to compare two contours. The hausdorff distance is defined as the "maximum distance of a set to the nearest point in the other set". Roughly speaking, it captures the maximum separation between the manual and the automatic contours. As observed in Figure 11 and Table 2 the experimental values obtained for the hausdorff distance correlate well with the error on the endpoints. This fact means that, most of the times, the major error on the automatic contour is located on the endpoint, with the shortest path algorithm recovering the true contour rapidly. A clear exception is patient #73, for which the error in the endpoints is negligible but the hausdorff distance between contours is very high. Here, the contour tracking algorithm missed to follow the contour, although it received proper endpoints.

The average distance between two contours captures better the perceived quality of the automatic breast contour. Here, the distance is averaged over the whole contour. Figure 12 and Table 3 summarize these results. As expected by visual inspection patient of #98 in Figure 10(c), although the error on the endpoint's location was high, the breast contour algo-

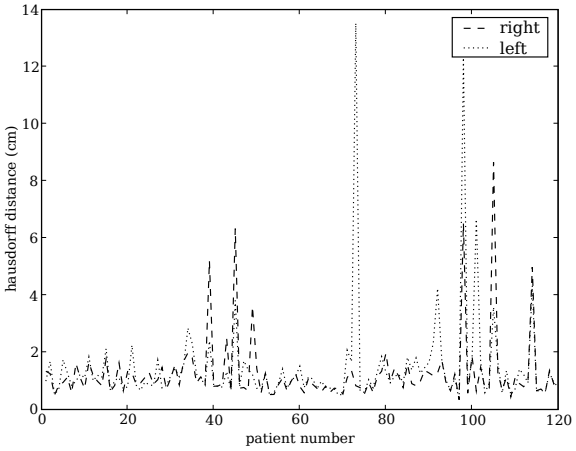


Figure 11: Evolution of hausdorff distance (cm) in the position of the breast contour over 120 photographs.

Table 2: Mean, standard deviation and maximum value of hausdorff distance in the position of the breast contour.

Hausdorff dist. (cm)	mean	std dev	max
left endpoint	1.5	1.7	13.5
right endpoint	1.3	1.2	8.7
total	1.4	1.5	13.5

Table 3: Mean, standard deviation and maximum value of average distance in the position of the breast contour.

Average dist. (cm)	mean	std dev	max
left endpoint	0.2	0.6	5.2
right endpoint	0.1	0.3	2.4
total	0.2	0.4	5.2

rithm recovered rapidly, translated into a small average distance (but a high hausdorff distance).

The consequences of the diverse errors enumerated before to the computer-aided medical system are different and have to be further studied. For example, one of the features used for the objective aesthetic evaluation of the BCCT is the difference between the levels of inferior breast contour points. This measure is quite robust over strong errors on the endpoints' position, as long as the breast contour is correctly tracked. Other features, such as the difference in the area of the breast are much more sensible to the positioning of the endpoints. One line of future investigation is the selection of features robust to the common errors of the automatic detection of notable points (contour endpoints, breast contour, nipples, etc) but still capturing adequately the aesthetic result, leading to a good classification performance.

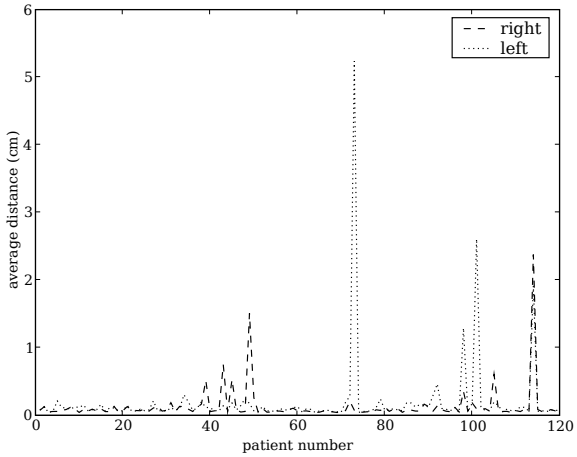


Figure 12: Evolution of the average distance (cm) in the position of the breast contour over 120 photographs.

5 CONCLUSIONS

A method has been described for applying graph concepts to the task of automatically extracting the breast contour in digital photographs of the torso of a patient, after submitted to a breast cancer conservative treatment. In the proposed framework the problem of finding the endpoints of the breast contour is formulated as a problem of finding strong contours between two regions, a concept introduced here for the first time. The breast contour is found as the solution to the shortest problem on a graph, after conveniently modelling the image as a weighted graph. Preliminary results indicate an excellent performance in the task of finding the external endpoints of the contour and a good performance on detecting the breast contour. Future work will focus on improvements to the algorithm including generalizing the solution to other typical patient positions in these studies.

ACKNOWLEDGEMENTS

This work was partially funded by Fundação para a Ciência e a Tecnologia (FCT) - Portugal through project PTDC/EIA/64914/2006.

REFERENCES

Al-Ghazal, S. K., Blamey, R. W., Stewart, J., and Morgan, A. L. (1999). The cosmetic outcome in early breast cancer treated with breast conservation. *European journal of surgical oncology*, 25:566–570.

Cardoso, J. S. and Cardoso, M. J. (2007a). Breast contour detection for the aesthetic evaluation of breast cancer conservative treatment. In *International Conference on Computer Recognition Systems (CORES'07)*.

Cardoso, J. S. and Cardoso, M. J. (2007b). Towards an intelligent medical system for the aesthetic evaluation of breast cancer conservative treatment. *Artificial Intelligence in Medicine*, 40:115–126.

Cardoso, M. J., Santos, A. C., Cardoso, J. S., Barros, H., and Oliveira, M. C. (2005). Choosing observers for evaluation of aesthetic results in breast cancer conservative treatment. *International Journal of Radiation Oncology, Biology and Physics*, 61:879–881.

Christie, D. R. H., O'Brien, M.-Y., Christie, J. A., Kron, T., Ferguson, S. A., Hamilton, C. S., and Denham, J. W. (1996). A comparison of methods of cosmetic assessment in breast conservation treatment. *Breast*, 5:358–367.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271.

Harris, J. R., Levene, M. B., Svensson, G., and Hellman, S. (1979). Analysis of cosmetic results following primary radiation therapy for stages i and ii carcinoma of the breast. *International Journal of Radiation Oncology Biology Physics*, 5:257–261.

Limbergen, E. V., Schueren, E. V., and Tongelen, K. . V. (1989). Cosmetic evaluation of breast conserving treatment for mammary cancer. 1. proposal of a quantitative scoring system. *Radiotherapy and oncology*, 16:159–167.

HUMAN FACE VERIFICATION BASED ON MULTIDIMENSIONAL POLYNOMIAL POWERS OF SIGMOID (PPS)

João Fernando Marar
Department of Computing, Adaptive Systems and Computational Intelligence Laboratory
Faculdade de Ciências, São Paulo State University, Bauru , São Paulo, Brazil
fermarar@fc.unesp.br, fermarar@di.fc.ul.pt

Helder Coelho
Department of Informatics, Laboratory of Agent Modelling
Faculdade de Ciências, Lisbon University, Lisbon, Portugal
hcoelho@di.fc.ul.pt

Keywords: Artificial Neural Network, Human Face Verification, Polynomial Powers of Sigmoid (PPS), Wavelets Functions, PPS-Wavelet Neural Networks, Activation Functions, Feedforward Networks.

Abstract: In this paper, we described how a multidimensional wavelet neural networks based on Polynomial Powers of Sigmoid (PPS) can be constructed, trained and applied in image processing tasks. In this sense, a novel and uniform framework for face verification is presented. The framework is based on a family of PPS wavelets,generated from linear combination of the sigmoid functions, and can be considered appearance based in that features are extracted from the face image. The feature vectors are then subjected to subspace projection of PPS-wavelet. The design of PPS-wavelet neural networks is also discussed, which is seldom reported in the literature. The Stirling Universitys face database were used to generate the results. Our method has achieved 92 % of correct detection and 5 % of false detection rate on the database.

1 INTRODUCTION

Systems based on biometric characteristics, such as face, fingerprints, geometry of the hands, iris pattern and others have been studied with attention. Face verification is a very important of these techniques because through it non-intrusive systems can be created, which means that people can be computationally identified without their knowledge. This way, computers can be an effective tool to search for missing children, suspects or people wanted by the law. Mathematically speaking, human face verification problem can be formulated as function approximation problems and from the viewpoint of artificial neural networks these can be seen as the problem of searching for a mapping that establishes a relationship from an input to an output space through a process of network learning.

Wavelet functions have been successfully used in many problems as the activation function of feedforward neural networks. An abundance of R&D has been produced on wavelet neural network area. Some successful algorithms and applications in wavelet neural network have been developed and reported in the literature (Pati and Krishnaprasad, 1993; Marar, 1997; Oussar and Dreyfus, 2000; Fan and Wang, 2005; Zhang and Pu, 2006; Avci, 2007; Jiang et al., 2007; Misra et al., 2007).

However, most of the aforementioned reports impose many restrictions in the classical backpropagation algorithm, such as low dimensionality, tensor product of wavelets, parameters initialization, and, in general, the output is one dimensional, etc.

In order to remove some of these restrictions, we develop a robust three layer PPS-wavelet multidimensional strongly similar to classical multilayer perceptron. The great advantage of this new approach is that PPS-Wavelets offers the possibility choice of the function that will be used in the hidden layer, without need to develop a new learning algorithm. This is a very interesting property for the design of new wavelet neural networks architectures. This paper is organized as follows. Section 2 introduces the wavelet sigmoidal function. Section 3 presents the framework used in this research. Section 4 deals with application of human face verification problem. Section 5 concludes this paper.

2 WAVELET FUNCTIONS

Two categories of wavelet functions, namely, orthogonal wavelets and wavelet frames (or non-orthogonal), were developed separately by different interests. An orthogonal basis is a family of wavelets that are linearly independent and mutually orthogonal, this eliminates the redundancy in the representation. However, orthogonal wavelets bases are difficult to construct because the wavelet family must satisfy stringent criteria (Daubechies, 1992). This way, for these difficulties, orthogonal wavelets is a serious drawback for their application to function approximation and process modeling (Oussar and Dreyfus, 2000). Conversely, wavelet frames are constructed by simple operations of translation and dilation of a single fixed function called the mother wavelet, which must satisfy conditions that are less stringent than orthogonality conditions.

Let $\varphi_j(x)$ a wavelet, the relation:

$$\varphi_j(x) = \varphi(d_j.(x-t_j))$$

where t_j is the translation factors and d_j is the dilation factors $\in \mathbf{R}$. The family of functions generated by $\mho$ can be defined as:

$$\mho = \left\{\varphi(d_j.(x-t_j)), t_j \text{ and } d_j \in \mathbf{R}\right\}$$

A family $\mho$ is said to be a frame of $L^2(\mathbf{R})$ if there exist two constants $c > 0$ and $C < \infty$ such that for any square integrable function f the following inequalities hold:

$$c\|f\|^2 \leq \sum_j |<\varphi_j, f>|^2 \leq C\|f\|^2$$

where $\varphi_j \in \mho$, $\|f\|$ denotes the norm of function f and $<\varphi_j, f>$ the inner product of functions. Families of wavelet frames of $L^2(\mathbf{R})$ are universal approximators (Pati and Krishnaprasad, 1993; Marar, 1997). In this work, we will show that wavelet frames allow practical implementation of multidimensional wavelets. This is important when considering problems of large input and output dimension. For the modeling of multi-variable processes, such as, the artificial neural networks biologically plausible, multidimensional wavelets must be defined. In the present work, we use multidimensional wavelets constructed as linear combination of sigmoid, denominated Polynomial Powers of Sigmoid Wavelet (PPS-wavelet).

2.1 Sigmoidal Wavelet Functions

In (Funahashi, 1989) is showed that:

Let $s(x)$ a function different of the constant function, limited and monotonically increase. For any $0 < \alpha < \infty$ the function created by the combination of sigmoid is described in Equation 1:

$$g(x) = s(x+\alpha) - s(x-\alpha) \tag{1}$$

where $g(x) \in L^1(\mathbf{R})$, i.e,

$$\int_{-\infty}^{\infty} g(x) < \infty$$

in particular, the sigmoid function satisfies this property.

Using the property came from the Equation 1, in (Pati and Krishnaprasad, 1993) boundary suggest the construction of wavelets based on addition and subtraction of translated sigmoidal, which denominates wavelets of sigmoid. In the same article show a process of construction of sigmoid wavelet by the substitution of the function $s(x)$ by $\Upsilon(qx)$ in the Equation 1. So, the Equation 2 is the wavelet function created in (Pati and Krishnaprasad, 1993).

$$\psi(x) = g(x+r) - g(x-r) \tag{2}$$

where $r > 0$. By terms of sigmoid function, the Equation 2, $\psi(x)$ is given by:

$$\psi(x) = \Upsilon(qx+a+r) - \Upsilon(qx-a+r) - \Upsilon(qx+a-r) + \Upsilon(qx-a-r) \tag{3}$$

where $q > 0$ is a constant that control the curve of the sigmoid function and α and $r \in \mathbf{R} > 0$.

Pati and Krishnaprasad demonstrated that the function $\psi(x)$ satisfies the admissibility condition for wavelets (Daubechies, 1992). The Fourier Transform of the function $\psi(x)$ is given by the Equation 4:

$$\int_{-\infty}^{\infty} \psi(x) e^{-iwx} dx = -i\frac{4\pi}{q} \frac{sin(w\alpha)sin(wr)}{sinh(\frac{\pi w}{q})} \tag{4}$$

In particular, we accepted for analysis and practical applications the family of sigmoid wavelet generated by the parameters $q = 2$ and $\alpha = r$, as example. So, the Equation 3 can be rewritten the following form:

$$\psi(x) = \Upsilon(2x+m) - 2\Upsilon(2x) - \Upsilon(2x-m) \tag{5}$$

where $m = \alpha + r$.

Following, partially, this research line, we present in the next section a technique for construction of wavelets based on linear combination of sigmoid powers.

3 POLYNOMIAL POWERS OF SIGMOID

The Polynomial Powers of Sigmoid (PPS) is a class of functions that have been used in recent years to solve a wide range of problems related to image and signal processing (Marar, 1997). Let $\Upsilon : \mathbf{R} \rightarrow$ **[0,1]** be a sigmoid function defined by $\Upsilon(x) = \frac{1}{1+e^{-x}}$. The $n^{th}-$power of the sigmoid function is a function $\Upsilon^n : \mathbf{R} \rightarrow$ **[0,1]** defined by $\Upsilon^n(x) = \left(\frac{1}{1+e^{-x}}\right)^n$.

Let Θ be set of all power functions defined by (6):

$$\Theta = \{\Upsilon^0(x), \Upsilon^1(x), \Upsilon^2(x), \ldots, \Upsilon^n(x), \ldots\} \quad (6)$$

An important aspect is that the power these functions, still keeps the form of the letter S. Looking the form created by the power functions of sigmoid, suppose that the n^{th} power of the sigmoid function to be represented by the following form:

$$\Upsilon^n(x) = \frac{1}{a_0 + a_1 e^{-x} + a_2 e^{-2x} + \cdots + a_n e^{-nx}} \quad (7)$$

where $a_0, a_{1,} a_{2,} \ldots, a_n$ are some integer values. The extension of the sigmoid power can be viewed like lines of a Pascal′s triangle. The set of function written by linear combination of polynomial powers of sigmoid is defined as PPS function. The degree of the PPS is given by the biggest power of the sigmoid terms.

3.1 Polynomial Wavelet Family on PPS

The derivative of a function $f(x)$ on $x = x_0$ is defined by:

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

since the limits there is. So, if we do the computation of the Equation 8 :

$$\frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \quad (8)$$

for a small value of Δx , showed have a good approximation for $f'(x_0)$. Naturally, Δx can be positive or negative. So, if is we use negative value for Δx, the expression will be:

$$\frac{f(x_0 - \Delta x) - f(x_0)}{-\Delta x} \quad (9)$$

This way, we can say that the arithmetic measure of the Equations 8 and 9 will be a good approximation for $f'(x_0)$ too. Then, we can write the following Equation 10:

$$f'(x_0) \simeq \frac{f(x_0 + \Delta x) - f(x_0 - \Delta x)}{2\Delta x} \quad (10)$$

By convenience, we consider $p = 2\Delta x$ and its substitution in the Equation 10. So, we have the Equation 11:

$$f'(x_0) \simeq \frac{f(x_0 + \frac{p}{2}) - f(x_0 - \frac{p}{2})}{p} \quad (11)$$

this point we computed an approximated value for the second derivative of $f(x)$ in $x = x_0$. From the Equation 11, changing $f(x)$ by $f'(x)$, we obtain the Equation 12 :

$$f''(x_0) \simeq \frac{f'(x_0 + \frac{p}{2}) - f'(x_0 - \frac{p}{2})}{p} \quad (12)$$

reusing the Equation 11, we can write:

$$f'(x_0 + \frac{p}{2}) \simeq \frac{f(x_0 + p) - f(x_0)}{p}$$

and

$$f'(x_0 - \frac{p}{2}) \simeq \frac{f(x_0) - f(x_0 - p)}{p}$$

using these results in the Equation12, we have an approximation of the second derivative of $f(x)$ in $x = x_0$ that is given by:

$$f''(x_0) \simeq \frac{f(x_0 + p) - 2f(x_0) + f(x_0 - p)}{p^2} \quad (13)$$

The approximation given by the Equation 13 is extremely adequate for the that $f(x)$ is a sigmoid function. Suppose that $f(x)$ is a sigmoid, for example, $\Upsilon(x)$. So, the second derivative of $\Upsilon(x)$ is approximated by the Equation 14:

$$\Upsilon''(x_0) \simeq \frac{\Upsilon(x_0 + p) - 2\Upsilon(x_0) + \Upsilon(x_0 - p)}{p^2} \quad (14)$$

Due the fact of the sigmoid function to be continuous and differentiable for any $x \in \mathbf{R}$, we can say that the Equation 14 is true for any x_0, then we can write the Equation 15, defined for all $x \in \mathbf{R}$.

$$\Upsilon''(x) \simeq \frac{\Upsilon(x_0 + p) - 2\Upsilon(x) + \Upsilon(x - p)}{p^2} \quad (15)$$

Comparison the Equations 15 and 5, we do there analysis for the approximation of the second derivative of sigmoid function. The first for values of $p \geq 1$ and the second for values of $p < 1$.

Case $p \geq 1$**:**

It is clear that the function given by the sigmoid second derivative approximation, Equation 15, also will have the same form of the Pati and Krishnaprasad functions, except of a p^2 constant that divides their amplitude. So, the following result is true: when $p > 1$ always there is a sigmoid wavelet which integral of the admissibility condition (Daubechies, 1992) limited the same integral of the Equation 15. Therefore, the approximation of the second derivative of the sigmoid function is a wavelet too.

Case $p < 1$:

In this case, we will analyze when p is going to zero, i.e.,

$$\lim_{p\to 0} \frac{\Upsilon'(x_0+p) - 2\Upsilon(x) + \Upsilon'(x-p)}{p^2} \qquad (16)$$

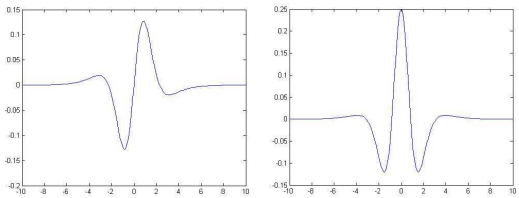
this limit tends to the second derivative of the function is given on PPS terms by:

$$\varphi_2(x) = 2\Upsilon(x)^3 - 3\Upsilon(x)^2 + \Upsilon(x) \qquad (17)$$

where we denominated $\varphi_2(x)$ the first wavelet the sigmoid function. The others derivatives, begin on the second, we considered true by derivative property by Fourier Transform (Marar, 1997). The successive derivation process of sigmoid functions, allowed to join a family of wavelets polynomial functions. Among many applications for this family of PPS-wavelets, special one is that those functions can be used like activation functions in artificial neurons. The following results correspond to the the analytical functions for the elements $\varphi_3(x)$ and $\varphi_4(x)$ that are represented by:

$$\varphi_3(x) = -6\Upsilon^4(x) + 12\Upsilon^3(x) - 7\Upsilon^2(x) + \Upsilon(x)$$

$$\varphi_4(x) = 24\Upsilon^5(x) - 60\Upsilon^4(x) + 50\Upsilon^3(x) - 15\Upsilon^2(x) + \Upsilon(x)$$



$\varphi_4(x)$ $\varphi_5(x)$

Figure 1: PPS-wavelets examples.

3.2 PPS Wavelet Neural Network

Let us consider the canonical structure of the multidimensional PPS-wavelet neural network (PPS-WNN), as shown in Figure 2.

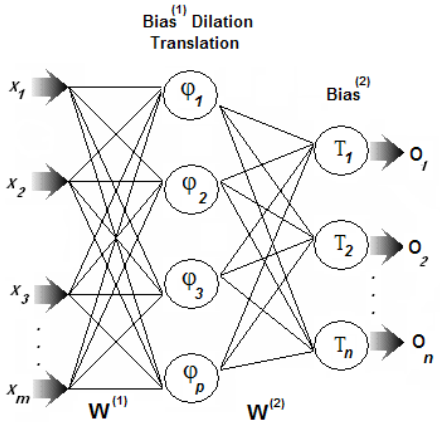


Figure 2: PPS-wavelet neural network Architectures.

For the PPS-WNN in Figure 2, when a input pattern $X = (x_1, x_2, \ldots, x_m)^T$ is applied at the input of the network, the output of the i^{th} neuron of output layer is represented as a function approximation problem, ie, $f : \mathbf{R}^m \rightarrow \mathbf{[0,1]}^n$, given by:

$$O_i(x) \simeq \Upsilon_i\left(\sum_{j=1}^{p} w_{ij}^{(2)} \varphi_j\left(d_j.\left(\sum_{k=1}^{m} w_{jk}^{(1)} x_k - b_j^{(1)}\right) - t_j\right) - b_i^{(2)}\right) \qquad (18)$$

where p is number of hidden neurons, $\Upsilon(.)$ is sigmoid function, $\varphi(.)$ is the PPS-wavelet, $w^{(2)}$ are weight between the hidden layer to the output layer, $w^{(1)}$ are weights between the input to the hidden layer, d are dilation factors and t are translation factors of the PPS-wavelet, $b^{(1)}$ and $b^{(2)}$ are bias factors of the hidden layer and output layer, respectively.

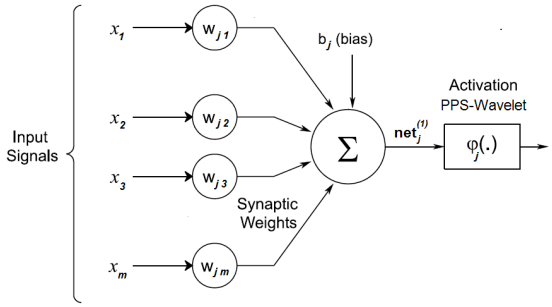


Figure 3: The Hidden Neuron of PPS-Wavelet Neural Network.

The PPS-WNN contains PPS-wavelets as the activation function in the hidden layer (Figure 3) and sigmoid function as the activation function in the output layer (Figure 4).

The output of the j^{th} PPS-wavelet hidden neuron (Figure 3) is given by :

$$\circledast_j = \varphi_j(d_j.(net_j^{(1)} - t_j))$$

where

$$net_j^{(1)} = \sum_{k=1}^{m} w_{jk}^{(1)} x_k - b_j^{(1)}$$

The output of the i^{th} output layer neuron (Figure 4)

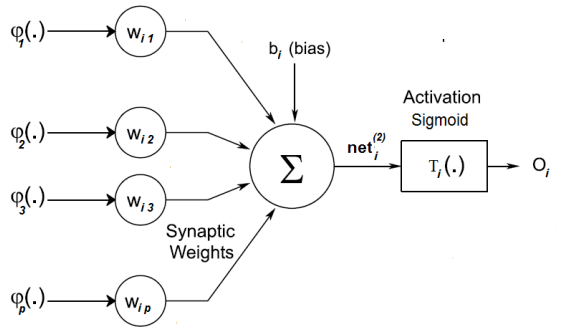


Figure 4: The Output Neuron of PPS-Wavelet Neural Network.

is given by:

$$\circledcirc_i = \frac{1}{1+exp(-net_i^{(2)})}$$

where

$$net_i^{(2)} = \sum_{j=1}^{p} w_{ij}^{(2)} \varphi_j(d_j.(net_j^{(1)} - t_j)) - b_i^{(2)}$$

The adaptive parameters of the PPS-WNN consist of all weights, bias, translations and dilation terms. The sole purpose of the training phase is to determine the "optimum" setting of the weights, bias, translations and dilation terms so as to minimize the difference between the network output and the target output. This difference is referred to as training error of the network. In the conventional backpropagation algorithm, the error function is defined as:

$$E = \frac{1}{2} \sum_{q=1}^{s} \sum_{i=1}^{n} (y_{qi} - o_{qi})^2 \quad (19)$$

where n is the dimension of output space, s is the number of training input patterns

The most popular and successful learning method for training the multilayer perceptrons is the backpropagation algorithm. The algorithm employs an iterative gradient descendent method of minimization which minimizes the mean squared error (L^2 norm) between the desired output (y_i) and network output (o_i). From Equations (18) and (19), we could deduce the partial derivatives of the error to each PPS-wavelet neural network parameter′s, which is given by:

Partial Equations of the Output Layer

$$\frac{\partial E}{\partial w_{ij}^{(2)}} = -\sum_{q=1}^{s} (y_{qi} - o_{qi}).o_{qi}.(1 - o_{qi}). \varphi_j(d_j.(net_{qj}^{(1)} - t_j)) \quad (20)$$

$$\frac{\partial E}{\partial b_i^{(2)}} = \sum_{q=1}^{s} (y_{qi} - o_{qi}).o_{qi}.(1 - o_{qi}) \quad (21)$$

Partial Equations of the Hidden Layer

$$\frac{\partial E}{\partial w_{jk}^{(1)}} = -d_j. \sum_{q=1}^{s} [\varphi'_j(d_j.(net_{qj}^{(1)} - t_j)).x_{qk}. \sum_{i=1}^{n} (y_{qi} - o_{qi}).o_{qi}.(1 - o_{qi}).w_{ij}^{(2)}] \quad (22)$$

$$\frac{\partial E}{\partial b_j^{(1)}} = \sum_{q=1}^{s} [\varphi'_j(d_j.(net_{qj}^{(1)} - t_j)).d_j. \sum_{i=1}^{n} (y_{qi} - o_{qi}).o_{qi}.(1 - o_{qi}).w_{ij}^{(2)}] \quad (23)$$

Partial Equations of the PPS-Wavelet Parameters

$$\frac{\partial E}{\partial d_j} = \sum_{q=1}^{s} \{[\varphi'_j(d_j.(net_{qj}^{(1)} - t_j)).(net_{qj}^{(1)} - t_j)]. \sum_{i=1}^{n} (y_{qi} - o_{qi}).o_{qi}.(1 - o_{qi}).w_{ij}^{(2)}\} \quad (24)$$

$$\frac{\partial E}{\partial t_j} = d_j \sum_{q=1}^{s} [\varphi'_j(d_j.(net_{qj}^{(1)} - t_j)). \sum_{i=1}^{n} (y_{qi} - o_{qi}).o_{qi}.(1 - o_{qi}).w_{ij}^{(2)}] \quad (25)$$

After computing all partial derivatives the network parameters are updated in the negative gradient direction. A learning constant γ defines the step length of the correction, r is the iteration and momentum factor is β. The corrections are given by:

$$w_{ij}^{(2)}(r+1) = w_{ij}^{(2)}(r) - \gamma.\frac{\partial E}{\partial w_{ij}^{(2)}} + \beta.(w_{ij}^{(2)}(r) - w_{ij}^{(2)}(r-1))$$

$$b_i^{(2)}(r+1) = b_i^{(2)}(r) - \gamma.\frac{\partial E}{\partial b_i^{(2)}} + \beta.(b_i^{(2)}(r) - b_i^{(2)}(r-1))$$

$$w_{jk}^{(1)}(r+1) = w_{jk}^{(1)}(r) - \gamma.\frac{\partial E}{\partial w_{jk}^{(1)}} + \beta.(w_{jk}^{(1)}(r) - w_{jk}^{(1)}(r-1))$$

$$b_j^{(1)}(r+1) = b_j^{(1)}(r) - \gamma.\frac{\partial E}{\partial b_j^{(1)}} + \beta.(b_j^{(1)}(r) - b_j^{(1)}(r-1))$$

$$d_j(r+1) = d_j(r) - \gamma.\frac{\partial E}{\partial d_j} + \beta.(d_j(r) - d_j(r-1))$$

$$t_j(r+1) = t_j(r) - \gamma.\frac{\partial E}{\partial t_j} + \beta.(t_j(r) - t_j(r-1))$$

4 HUMAN FACE VERIFICATION

This study presents a system for detection and extraction of faces based on the approach presented in (Lin and Fan, 2001), which consists of finding isosceles triangles in an image, as the mouth and eyes form that geometric figure when linked by lines. In order for these regions to be determined, the images must be converted into binary images, thus the vertices of the triangles must be found and a rectangle must be cut out around them so that their size can be brought to normal and the area can be fed into a second part of the system that will analyze whether or not it is a real face. three different approaches are tested here: A weighing mask is used to score the region, proposed by Lin and Fan (Lin and Fan, 2001), a classical MLP backpropagation (MLP-BP) and PPS-wavelet neural network, for the analysis to be performed.

4.1 Image Treatment

First the image was read with the purpose of allocating a matrix in which each cell indicates the level of brightness of the correspondent pixel; then, it is converted into a binary matrix by means of a Threshold parameter T, because the objects of interest in our case are darker than the background. This stage changes to 1 (white) a brightness level greater than T and to 0 (black). In most of the cases, due to noise and distortion in the input image, the result of the binary transformation can bring a partition image and isolated pixels. Morphologic operations - opening followed by closing - are applied with the purpose of solving or minimizing this problem (Gonzalez and Woods, 2002). The Figure 5 shows the result of these operations.

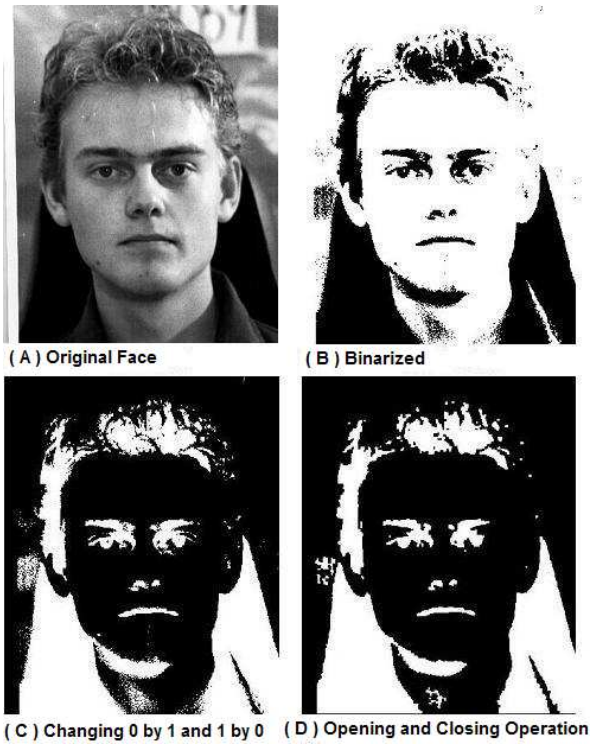


Figure 5: Image treatment after morphologic operations.

✓Segmentation of Potential Face Regions

After binarization the task is finding the center of three 4-connected components that meet the following characteristics:

1. vertex of an isosceles triangle (Lin and Fan, 2001);
2. the Euclidean distance between the eyes must be 90-100 % the distance between the mouth and the central point between the eyes (Lin and Fan, 2001);
3. the triangle base is at the top of the image.

The last restriction does not allow finding upside down faces, but it significantly reduces the number of triangles in each image, thus reducing the processing time to the following stages. For example, the numbers of triangles found in Figure 5(D), with this restriction 399 and without restriction 769.

The opening and closing operations are vital, since it is impossible to determine the triangles without this image treatment. The processing mean time

to find the results presented was 4 seconds; on the other hand, 8 hours were insufficient in an attempt at finding the same results using a Pentium 4 with 2.4 Ghz processor in Figure 5(C).

✓**Normalization of Potential facial Regions**

Once the potential face regions that we have selected in the previous section are allowed to have different sizes. All regions had to be normalized to the $(60x60)$ pixels size by bi-cubic interpolation technique, because every potential regions needs to present the same amount of information for comparison. So, normalization of a potential region can reduce the effects of variation in distance and location.

4.2 Face's Pattern Recognition

The purpose of this stage is to decide whether a potential face region in an image (the region extracted in the first part of the process) actually contains a face. To perform this verification, two methods were applied : The weighting mask function, described by Lin and Fun (Lin and Fan, 2001) and PPS-wavelet neural network.

✓**The Weighting Mask Function**

The function Weighting Mask, according to the author, it is based on the following idea: If the normalized potential region is really contains a face, it should have high similarity to the mask that is formed by 10 binary training faces (Mask Generation). Every normalized potential facial region is applied into the weighting mask function that is used to compute the similarity between the normalized potential facial region and the mask. The computed value can be used in deciding whether a potential region contains a face or not.

Mask Generation

The mask was created using 10 images. The first five are pictures of females and the others are pictures of males. All of them were manually segmented, binarized, normalized, morphologically treated (opening and closing) and then the sum of the correspondent cell of each image was stored in the 11^{th} matrix. Finally, that matrix was binarized with another Threshold T, for which values lower than or equal to T were replaced by 0, and the others by 1. The result was improved with T=4. Whereas at lower values the areas of the eyes and mouth become too big, at higher values these areas almost disappear. In both cases, determining the triangles is considerably difficult.

Weighting Mask Algorithm

The algorithm used to decide whether a potential face (R) contains a real face is based on the idea that the binary image of a face is highly similar to that of the mask.

Begin
Input the region R and mask M; p=0;
For all pixels of R and M
IF the pixel from R and M is white
Then p = p+6;
IF the pixel from R and M is black
Then p = p+2;
IF the pixel from R is white and that from M is black
Then p = p - 4;
IF the pixel from R is black and that from M is white
Then p = p - 2;
End

A set experimental results demonstrates that the threshold values should be a set between face $3400 <= p >= 6800$ (Marar et al., 2004).

✓**PPS-wavelet Neural Network**

In order to demonstrate the efficiency of the proposed model. Two PPS-WNNs, one with the activation function $\varphi_2(.)$ and the other with $\varphi_5(.)$ in the hidden layer, were implemented to analyze when a potential face region really contains a face. However, the raw data face, (60 x 60) pixels, cannot be used directly for the training the networks because the features are deep hidden. Therefore, we used the Principal Components Analysis (PCA) method to create a face space that represents all the faces using a small set of components (Marar, 1997). For this purpose we consider the first 15 components as the extracted features or face space. In that case study, 100 manually segmented faces (50 women and 50 men) and more 40 non-face random images were used to network training.

Therefore, the PPS-WNNs and classical MLP-BP architectures with 15 units in the input layer, with 16 PPS-wavelet neurons in the hidden layer and with 2 neurons in the output layer were designed and trained. Here, in the output layer, we represented face by the vector (1,0) and non-face by the vector (0,1). We used, as test, the same regions (R) applied to the previous method.

5 RESULTS

Several tests were performed to determine an ideal threshold value for the conversion of the images into binary figures. In a scale from 0 (black) to 1 (white), 0.38 was empirically determined as a good value to most of the images, but to darker images 0.22 was a better value. The test was done through the use of 100 images (50 male and 50 female) with two different threshold values from (Department, 2003). The results are shown in Table 1.

Table 1: Face verification results with 2 threshold values.

	Threshold	0.22	0.38
Weighting Mask	Correct Detection	81 %	48 %
	False Detection	25 %	21 %
Classical MLP-BP	Correct Detection	83 %	35 %
	False Detection	28 %	17 %
PPS-WNN $\varphi_2(.)$	Correct Detection	85%	63 %
	False Detection	15 %	23 %
PPS-WNN $\varphi_5(.)$	Correct Detection	92 %	51 %
	False Detection	5 %	11 %

The best result for T=0.22 is explained by the low brightness and consequently low contrast of the images in the set. All the images used are at an 8 bit gray scale and 540 x 640 pixels. All tests were performed in an IBM -compatible PC, Pentium 4 with 2.4 Ghz processor, 1Gb RAM memory.

6 CONCLUSIONS

The face recognition is an active research area for security. However, it is still a complex and challenging research topic because the human face may change its appearance due to the internal variations such as facial expressions, beards, mustaches, hair styles, glasses, ageing, surgery and the external distortions such as scale, lighting, position and face occlusion. In this paper, we showed the basic concepts and technics of Polynomial Powers of Sigmoid and how to build multidimensional wavelet neural networks starting from this definition. We chose this application due to the complexity of image processing problems. The obtained results suppose to validate the new method for new and future applications in the artificial intelligence area.

ACKNOWLEDGEMENTS

We would like to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) process number $3634/06-0$ and the Lisbon University that supported this investigation.

REFERENCES

Avci, E. (2007). An expert system based on wavelet neural network-adaptive norm entropy for scale invariant texture classification. *Expert Systems with Applications*, 32:919–926.

Daubechies, I. (1992). *Ten lecture on wavelets*. Society for Industrial and Applied Mathematics (SIAM).

Department, P. (2003). Psycological image collection at stirling university. http://pics.psych.stir.ac.uk.

Fan, J. and Wang, X. F. (2005). A wavelet view of small-world networks. *IEEE Transactions on Circuits and Systems*, pages 1–4.

Funahashi, K. (1989). On the approximate realization of continuos mappings by neural networks. *Neural Networks*, (2):183–192.

Gonzalez, R. C. and Woods, R. E. (2002). *Digital Image Processing*. Prentice-Hall, Inc.

Jiang, X., Mahadevan1, S., and Adeli, H. (2007). Bayesian wavelet packet denoising for structural system identification. *Struct. Control Health Monit.*, 14:333–356.

Lin, C. and Fan, K.-C. (2001). Triangle basead approuch to detection of human face. *Pattern Recognition Society*, pages 941–944.

Marar, J. F. (1997). *Polinomios Potências de Sigmoide (PPS): Uma nova Técnica para Aproximação de Funções, Construção de Wavenets e suas aplicações em Processamento de Imagens e Sinais*. PhD thesis, Universidade Federal de Pernambuco - Departamento de Informática.

Marar, J. F., Costa, D., Pinheiro, O., and Filho, E. (2004). Adaptative techniques for the human faces detection. In *6th International Conference on Enterprise Information Systems*, volume 2, pages 465–468.

Misra, B. B., Dash, P. K., and Panda, G. (2007). Pattern classification using local linear wavelet neural network. *International Conference on Information and Communication Technology*, pages 92–95.

Oussar, Y. and Dreyfus, G. (2000). Initialization by selection for wavelet neural traing. *Neurocomputing*, 34:131–143.

Pati, Y. and Krishnaprasad, P. (1993). Analysis and synthesis of feedforward neural networks using discrete affine wavelet transformations. *IEEE Transactions on Neural Networks*, 4(1):73–85.

Zhang, H. and Pu, J. (2006). A novel self-adaptive control framework via wavelet neural netwok. In *6th World congress on intelligent control and automation*, pages 2254–2258.

GESTURE THERAPY
A Low-Cost Vision-Based System for Rehabilitation after Stroke

L. Enrique Sucar[1], Ron S. Leder[2], David Reinkensmeyer[3], Jorge Hernández[4], Gidardo Azcárate[5]
Nallely Casteñeda[6] and Pedro Saucedo[7]
[1]*Departamento de Computación, INAOE, Tonantzintla, Puebla, México*
esucar@inaoep.mx
[2]*División de Ingeniería Eléctrica, UNAM, México D.F., México*
rleder@ieee.org
[3]*Department of Mechanical and Aerospace Engineering, UC Irvine, USA*
dreinken@uci.edu
[4]*Unidad de Rehabilitación, INNN, Mexico D.F., Mexico*
jhfranco@medicapolanco.com
[5]*ITESM Campus Cuernavaca, Morelos, México*
A01125613@itesm.mx
[6]*Unidad de Rehabilitación, INNN, Mexico D. F., Mexico*
alvacasana@yahoo.com.mx
[7]*Universidad Anáhuac del Sur, Mexico D.F., Mexico*
saucedouribe@gmail.com

Keywords: Rehabilitation, stroke, therapeutic technology.

Abstract: An important goal for rehabilitation engineering is to develop technology that allows individuals with stroke to practice intensive movement training without the expense of an always-present therapist. We have developed a low-cost, computer vision system that allows individuals with stroke to practice arm movement exercises at home or at the clinic, with periodic interactions with a therapist. The system intgrates a web-based system for facilitating repetitive movement training, with state-of-the art computer vision algorithms that track the hand of a patient and obtain its 3-D coordinates, using two inexpensive cameras and a conventional personal computer. An initial prototype of the system has been evaluated in a pilot clinical study with positive results.

1 INTRODUCTION

Each year in the U.S. alone over 600,000 people survive a stroke (ASA 2004), and similar figures exist in other countries. Approximately 80% of acute stroke survivors lose arm and hand movement skills. Movement impairments after stroke are typically treated with intensive, hands-on physical and occupational therapy for several weeks after the initial injury. Unfortunately, due to economic pressures on health care providers, stroke patients are receiving less therapy and going home sooner. The ensuing home rehabilitation is often self-directed with little professional or quantitative feedback. Even as formal therapy declines, a growing body of evidence suggests that both acute and chronic stroke survivors can improve movement ability with intensive, supervised training. Thus, an important goal for rehabilitation engineering is to develop technology that allows individuals with stroke to practice intensive movement training without the expense of an always-present therapist.

We have developed a prototype of a low-cost, computer vision system that allows individuals with stroke to practice arm movement exercises at home or at the clinic, with periodic interactions with a therapist. The system makes use of our previous work on a low-cost, highly accessible, web-based system for facilitating repetitive movement training, called "Java Therapy", which has evolved into T-WREX (Fig. 1) (Reinkensmeyer 2002 and Sanchez 2006). T-WREX provides simulation activities relevant to daily life. The initial version of Java Therapy allowed users to log into a Web site, perform a customized program of therapeutic activities using a mouse or a joystick, and receive

quantitative feedback of their progress. In preliminary studies of the system, we found that stroke subjects responded enthusiastically to the quantitative feedback provided by the system. The use of a standard mouse or joystick as the input device also limited the functional relevance of the system. We have developed an improved input device that consists of an instrumented, anti-gravity orthosis that allows assisted arm movement across a large workspace. However, this orthosis costs about $4000 to manufacture, limiting its accessibility. Using computer vision this system becomes extremely attractive because it can be implemented with low cost (i.e. using an inexpensive camera and conventional computer).

For "Gesture Therapy" we combine T-WREX with state-of-the art computer vision algorithms that track the hand of a patient and obtain its 3-D coordinates, using two inexpensive cameras (web cams) and a conventional personal computer (Fig. 2). The vision algorithms locate and track the hand of the patient using color and motion information, and the views obtained from the two cameras are combined to estimate the position of the hand in 3-D space. The coordinates of the hand (X, Y, Z) are sent to T-WREX so that the patient interacts with a virtual environment by moving his/her impaired arm, performing different tasks designed to mimic real life situations and thus oriented for rehabilitation. In this way we have a low-cost system which increases the motivation of stroke subjects to follow their rehabilitation program, and with which they can continue their arm exercises at home.

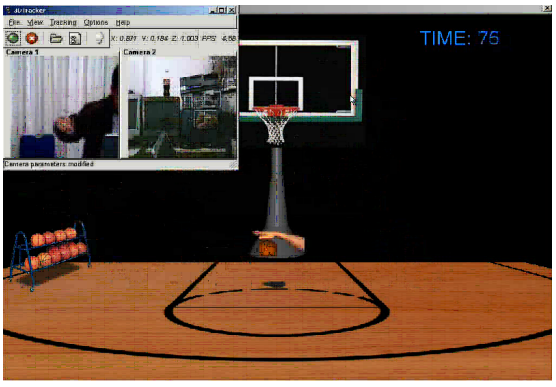


Figure 1: Screen shot of T-WREX. Arm and hand movements are focused as a mouse pointer to activate an object in the simulation. In this case a hand interacts with a basketball. The upper left insert shows the camera views, frontal and side, of the patient's hand tracked by the system.

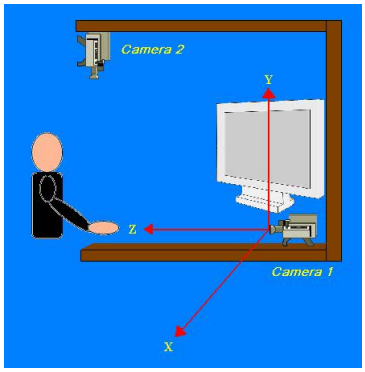


Figure 2: Set up for the Gesture Therapy system. The patient is seated in front of a table that serves as a support for the impaired arm, and its movements are followed by two cameras. The patient watches in a monitor the simulated environment and his/her control of the simulated actuator.

A prototype of this system has been installed at the rehabilitation unit at the National Institute of Neurology and Neurosurgery (INNN) in Mexico City, and a pilot study was conducted with a patient diagnosed with ischemic stroke, left hemi paresis, with a time of evolution of 4 years. After 6 sessions with Gesture Therapy, the results based on the therapist and patient opinions are positive, although a more extensive controlled clinical trial is required to evaluate the impact of the system in stroke rehabilitation. In this paper we describe the Gesture Therapy system and present the results of the pilot clinical study.

2 METHODOLOGY

Gesture Therapy integrates a simulated environment for rehabilitation (Java Therapy) with a gesture tracking software in a low-cost system for rehabilitation after stroke. Next we describe each of these components.

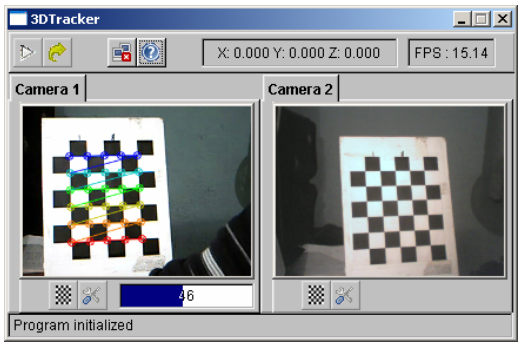


Figure 3: Reference pattern used for obtaining the intrinsic parameters of each camera (camera 1 and 2).

2.1 Java Therapy/T-WREX

The Java Therapy/T-WREX web-based user interface has three key elements: therapy activities that guide movement exercise and measure movement recovery, progress charts that inform users of their rehabilitation progress, and a therapist page that allows rehabilitation programs to be prescribed and monitored

The therapy activities are presented in the software simulation like games and the system configuration allows therapists to customize the software to enhance the therapeutic benefits for each patient, by selecting a specific therapy activity among others in the system

The therapy activities were designed to be intuitive even for patients with minimal cognitive or perceptual problems to understand. These activities are for repetitive daily task-specific practice and were selected by its functional relevance and inherent motivation like grocery shopping, car driving, playing basketball, self feeding, etc.

Additionally, the system gives objective visual feedback of patient task performance, and patient progress can be illustrated easily by the therapist by a simple statistical chart. The visual feedback has the effect of enhancing motivation and endurance along the rehabilitation process by patients awareness of his/her progress.

2.2 Gesture Tracking

Using two cameras (stereo system) and a computer, the hand of the user is detected and tracked in a sequence of images to obtain its 3-D coordinates in each frame, which are sent to the T-WREX environment. This process involves several stages:

- Calibration,
- Segmentation,
- Tracking,
- 3-D reconstruction.

Next we describe each stage.

2.2.1 Calibration

To have a precise estimation of the 3-D position in space of the hand, the camera system has to be calibrated. The calibration consists in obtaining the intrinsic (focal length, pixel size) and extrinsic (position and orientation) parameters of the cameras. The intrinsic parameters are obtained via a reference pattern (checker board) that is put in front of each camera, as shown in figure 3.

The extrinsic parameters are obtained by giving the system the position and orientation of each camera in space with respect to a reference point, see figure 2. The reference point could be the lens of one of the cameras, or an external point such as a corner of the table. The colors on the checker board pattern and the status bar shown in figure 3 above indicate the progress of the calibration process.

Note that the calibration procedure is done only once and stored in the system, so in subsequent sessions this procedure does not need to be repeated, unless the cameras are moved or changed for other models.

2.2.2 Segmentation

The hand of the patient is localized and segmented in the initial image combining color and motion information. Skin color is a good clue to point potential regions where there is a hand/face of a person. We trained a Bayesian classifier with many (thousand) samples of skin pixels in HSV (hue, saturation, value), which is used to detect skin pixels in the image. Additionally, we use motion information based on image subtraction to detect moving objects in the images, assuming that the patient will be moving his impaired arm. Regions that satisfy both criteria, skin color and motion, are extracted by an intersection operation, and this region corresponds to the hand of the person. This segment is used as the initial position of the hand for tracking it in the image sequence, as described in the next section. This procedure is applied to both images, as illustrated in figure 4.

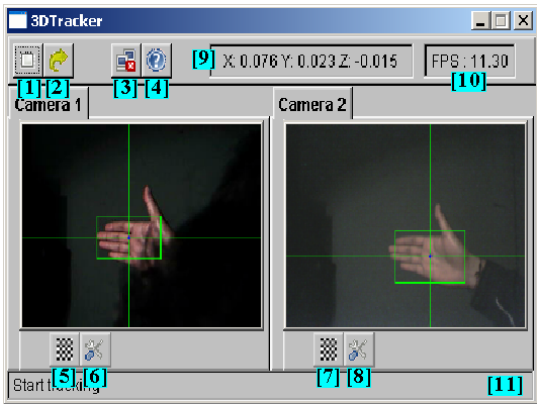


Figure 4: Hand detection and segmentation in both images. The approx. hand region is shown as a rectangle, in which the center point is highlighted, used later for finding the 3-D coordinates.

The system can be confused with objects that have a similar color as human skin (i.e wood), so we assume that this does not occur. For this it is recommended that the patient uses long sleeves, and to cover the table and back wall with a uniform cloth

in a distinctive color (like black or blue). It is also recommended that the system is used indoors with artificial lighting (white). Under these conditions that system can localize and track the hand quite robustly in real time.

2.2.3 Tracking

Hand tracking is based on the *Camshift* algorithm (Bradski, 1998). This algorithm uses only color information to track an object in an image sequence. Based on an initial object window, obtained in the previous stage, Camshift builds a color histogram of the object of interest, in this case the hand. Using a search window (define heuristically according to the size of the initial hand region) and the histogram, Camshift obtains a probability of each pixel in the search region to be part of the object, and the center of the region is the "mean" of this distribution. The distribution is updated in each image, so the algorithm can tolerate small variation in illumination conditions.

In this way, the 2-D position of the hand in each image in the video sequence is obtained, which corresponds to the center point of the color distribution obtained with Camshift. The 3-D coordinates are obtained by combining both views, as described in the next section.

2.2.4 3-D Reconstruction

Based on the 2-D coordinates of the center point of the image region in each image, the 3-D coordinates are obtained in the following way. For each image, a line in 3-D space is constructed by connecting the center of the hand region and the center of the camera lens, based on the camera parameters. This is depicted in figure 5. Once the two projection lines are obtained, their intersection provides the coordinates in 3-D (*X, Y, Z*).

Thus, we have the 3-D position of the hand for each processed image pair (about 15 frames per second in a standard PC), which are sent to T-WREX so that the patient can interact with the virtual environments.

3 PILOT STUDY

We performed a pilot study with one patient using "Gesture Therapy" at the National Institute for Neurology and Neurosurgery (INNN) in Mexico City. The purpose of this pilot study was to improve the protocol for a larger clinical trail with Gesture Therapy, anticipating potential problems and gaining experience using the technology in the hospital setting.

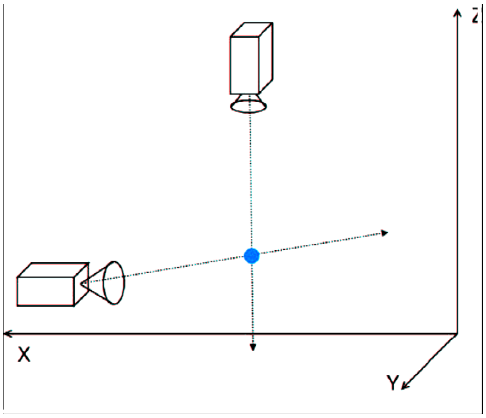


Figure 5: Estimation of the 3-D position of the hand by intersecting the projection lines obtained from the images.

The patient was diagnosed with ischemic stroke, left hemi paresis, with a time of evolution of 4 years. An evaluation with the Fugl-Meyer (Fugl-Meyer, 1975) scale was performed at the start and end of the study.

The patient used Gesture Therapy for 6 sessions, between 20 and 45 minutes each session. The main objective of the exercises was the control of the distal portion of the upper extremity and hand. The patient performed pre exercises for stretching, relaxation, and contraction of the fingers and wrist flexors and extensors. The patient performed several of the simulated exercises in the virtual environment, increasing in difficulty as the sessions progressed (clean stove, clean windows, basketball, paint room, car race).

After the 6 sessions the patient increased his capacity to voluntarily extend and flex the wrist through relaxation of the extensor muscles. He also tried to do bimanual activities (such as take and throw a basket ball) even if he maintained the affected left hand closed; he increased use of the affected extremity to close doors.

In the therapist's opinion: "The GT system favours the movement of the upper extremity by the patient. It makes the patient maintain the control of his extremity even if he does not perceive it. GT maintains the motivation of the patient as he tries to perform the activity better each time (more control in positioning the extremity, more speed to do the task, more precision). This particular patient gained some degree of range of movement of his wrist. There are still many problems with the fingers flexor synergy, but he feels well and motivated with his achievements. It is also important to note the motivation effect the system has on patient endurance to complete the treatment until the last

day by increasing the enthusiasm of the patient in executing the variety of rehabilitation exercises."

In the patient's opinion: "At the beginning I felt that my arm was too "heavy", and at the shoulder I felt as if there was something cutting me, now I feel it less heavy and the cutting sensation has also been reduced."

An "analogical visual scale" in the range 1-10 (very bad, ..., excellent) was applied, asking the patient about the treatment based on GT, he gave it a 10. Asked about if he will like to continue using GT, his answer was "YES".

The Fugl-Meyer scale (3 points increase) was not sufficiently sensitive to detect the clear clinical and subjective improvement in the patient.

4 CONCLUSIONS AND FUTURE WORK

This single case shows the importance of motivation in rehabilitation. Involving the patient in simulated daily activities helps the psychological rehabilitation component as well. The potential ease of use, motivation promoting characteristics, and objective quantitative potential are obvious advantages to this system. The patient can work independently with reduced therapist interaction. With current technology the system can be adapted to a portable low-cost device for the home including communications for remote interaction with a therapist and medical team.

It is possible to extend the system to a full arm tracking, including wrist, hand and fingers for more accurate movements. Movement trajectories can be compared and used to add a new metric of patient progress. To make the system easier to use a GUI tool is planned for system parameters configuration, including the camera. Future work includes more games to increase the variety of therapy solutions and adaptability to patient abilities, so that a therapist or patient can match the amount of challenge necessary to keep the rehabilitation advancing.

In the current low-cost, vision-based system the table top serves as an arm support for 2D movement until the patients are strong enough to lift their arms into 3D. Extending the system to wrist, hand, and finger movement is planned to make a full superior extremity rehabilitation system.

Wrist accelerometers can be used to increase the objectivity of clinical studies in addition to subjective reports of patients and caregivers; especially when the patient spends less time in the clinic. (Uswatte 2006). fMRI of patients' brains, pre and post training, are planned for increasing our understanding of the biological basis for rehabilitation (Johansen-Berg 2002).

ACKNOWLEDGEMENTS

This work was supported in part by a grant from UC-MEXUS/CONACYT.

REFERENCES

American Stroke Association, 2004. Retrieved July 10, 2007 from http://www.strokeassociation.org.

G.R. Bradski, 1998. Computer vision face tracking as a component of a perceptual user interface. In *Workshop on Applications of Computer Vision,* pp. 214-219.

A. R. Fugl-Meyer, L. Jaasko, I. Leyman, S. Olsson, and S. Steglind, 1975. The post-stroke hemiplegic patient: a method for evaluation of physical performance. In *Scand J Rehabil Med,* vol. 7, pp. 13-31.

D. Reinkensmeyer, C. Pang, J. Nessler, and C. Painter, 2002. Web-based telerehabilitation for the upper-extremity after stroke. In *IEEE Transactions on Neural Science and Rehabilitation Engineering,* vol. 10, pp. 1-7.

R. J. Sanchez, J. Liu, S. Rao, P. Shah, R. Smith, T. Rahman, S. C. Cramer, J. E. Bobrow, and D. Reinkensmeyer, 2006. Automating arm movement training following severe stroke: Functional exercise with quantitative feedback in a gravity-reduced environment. In *IEEE Trans. Neural. Sci. Rehabil. Eng*., vol. 14(3), pp. 378-389.D. Reinkensmeyer, S. Housman, Vu Le, T. Rahman and R. Sanchez, 2007. Arm-Training with T-WREX After Chronic Stroke: Preliminary Results of a Randomized Controlled Trial. In *ICORR 2007, 10th International Conference on Rehabilitation Robotics,* Noordwijk.

D. Reinkensmeyer and S. Housman, 2007. If I can't do it once, why do it a hundred times?: Connecting volition to movement success in a virtual environment *motivates people to exercise the arm after stroke.* In *IWVR 2007*

G, Uswatte, C. Giuliani, C. Winstein, A. Zeringue, L. Hobbs, SL Wolf, 2006. Validity of accelerometry for monitoring real-world arm activity in patients with subacute stroke: evidence from the extremity constraint-induced therapy evaluation trial. *Arch Med Rehabil:* 86:1340-1345.

H. Johansen-Berg, H. Dawes, C. Guy, S.M. Smith, D.T. Wade and P.M. Matthews, 2002 Correlation between motor improvements and altered fMRI activity after rehabilitative therapy. *Brain Journal:* 125, 2731-2742

CLINICAL PRACTICAL GUIDELINE EDITOR
Clinical Practical Guideline-based Decision Support Tool

Plavčak Aleš
MIPS programska oprema d.o.o., Cesta k Tamu 7, 2000 Maribor, Slovenia
ales.plavcak@gmail.com

Keywords: Decision-support systems, medical informatics, clinical practice guidelines, knowledge modelling.

Abstract: According to quick growth of information and complexity of medicine, the development of informatics in medicine is in full bloom. Medical decision support systems have been developed to help choose the appropriate medical treatment procedures, ensure the quality of health care and enable the control of resources. Clinical Practical Guidelines have greatly contributed to the accelerated development. Many different modelling methods and tools have been developed for executing guidelines. Here, the three of many applicable guideline modelling techniques are represented in greater scope. Also, a new technique of representing the clinical knowledge has been designed, taken from the studies of already existing models. The implementation of the application for editing, browsing and executing clinical guidelines has been implemented as well. The application is capable of generating recommendations for a specific clinical state and evaluation of the already existing health care process. This paper covers the general presentation of informatics in medicine and the techniques for modelling of medical knowledge which nowadays represents a gateway for prosperous development, and paves the way for broader use and implementation.

1 INTRODUCTION

The experiences of individual physicians, opinions of professors in medical schools, medical textbooks, clinical journals and clinical trials have guided the practice of medicine for most of this century, so the standardization is not necessarily obligatory, but strongly recommendable. The medical community has always standardized medical care to some degree in order to provide what it thought was the best care, to efficiently use resources, to satisfy patients, and to withstand third party scrutiny.

The idea of studying which treatments work best is nothing new, although systematic treatment procedures are a novelty in many branches of medicine. It is well known what the right things to be done are, but we have to make it happen, so here is methodology and computer based decision support systems to assist practitioners and patients in making decisions about appropriate management of specific clinical conditions.

Our use of computers has been driven not only by the increasing need to manage large amounts of information, but also by the imperative to make evidence based and cost effective decisions on a daily basis. Furthermore, there is accumulating evidence to prove that computer aided medical tools address the growing information needs of the busy clinician and improve healthcare processes as well as patient outcomes. In turn, this has led to the rapid proliferation of a variety of clinical decision support system (CDSS). A computerised CDSS is a computer based tool using explicit knowledge to generate patient specific advice or interpretation. It is now universally agreed that conforming to state-of-the-art guidelines are the best way to improve the quality of CDSS.

Nowadays the basic of developing a CDSS is a clinical practice guideline (CPG) that is a subject discussed by number of researchers who are trying to develop different technologies for delivering computerized guidelines in clinical care. The new research movement could revolutionise the health care industry by improving quality and reducing costs, say the experts that are studying, developing and evaluating CPGs. Above is depicted a newly proposed design and implementation of application for modelling, executing and evaluating CPG.

2 KNOWLEDGE MODELLING

Medical experts are increasingly expected to always make the best decision. This is difficult. The amount of medical information in the world is increasing, yet then capacity of a human brain is not. Computers have the ability to help deal with all this information, so generally speaking, computers are better than humans at managing loads of information and solving complex problems.

To be effective, these tools should have access to the patient's medical record, use standard medical vocabularies, have clear semantics, and facilitate knowledge maintenance and sharing. In addition to that they need to be sufficiently expressive in order to explicitly capture the design of the rational process and outcome intentions of the guideline's author, while at the same time they must enable the attending physician to use their own preferred methods to achieve a certain degree of flexibility during the application runtime.

The level of standardization in today's guidelines is relatively new. When predicated on sound medical and scientific data, these guidelines can lessen provider variability in treatment and diagnosis. Better standardization allows better measurement of resources used and assessment of benefits obtained. These guidelines can be particularly effective when applied to high-prevalence, high-cost diseases or conditions.

2.1 Clinical Practice Guidelines

CPGs are developed to reduce inappropriate variations in practice, to improve health care quality, and to help control costs. Although the importance of guidelines is widely recognized, health care organizations typically pay more attention to guideline development than to guideline implementation for routine use during care process.

The American Medical Association calls them "practice parameters" and defines them as "...strategies for patient management developed to assist physicians in clinical decision making." As already mentioned, they should not be rigid and static; rather, they should be flexible and dynamic road maps aimed at reducing clinically significant and unexplained variations in patient care process.

"CPG are systematically developed statements, based on best evidence, intended to assist practitioners and patients in making decisions about appropriate management of specific clinical conditions" (Institute of Medicine, 1990). This definition emphasizes the decision-making aspect of clinical practice guidelines.

Samson Tu and Mark Musen have identified five principle tasks that computerised guidelines and guideline representation methods should be capable of supporting: making decisions, sequencing actions, setting goals (e.g. specific patient states) to be achieved, interpreting data, refining actions (i.e. breaking up into sub-components).

But not all of them are able to fulfil all the principles. They cover many methods and demonstrate the use of different representation formalisms and computational techniques.

2.2 Guideline Modelling Methods

Based on a literature search of computer-based guideline specific representation models, three published research projects were included in this review and represent the base for developing a CPG support tool. Arden Syntax was chosen for its simplicity and represents a pioneering achievement for guidelines. PROforma is an easy understandable and flexible language for encoding medical knowledge. GLIF represents a complex guideline modelling method and has well defined object-orientated design.

2.2.1 Arden Syntax

Arden Syntax is a standard, formal procedural language that represents medical algorithms in clinical information systems as Medical Logic Modules (MLMs), and uses rule-based specification for encoding medical knowledge. It is the first standard for representing medical knowledge. An MLM is a hybrid between a production rule (i.e. an "if-then" rule) and a procedural formalism. Each MLM is invoked as if it were a single-step "if-then" rule, but then it executes serially as a sequence of instructions, including queries, calculations, logic statements and write statements.

Arden was developed for embedding MLMs into proprietary clinical information systems. It was specially designed to support clinical decision making. An individual MLM should contain sufficient logic to make a single medical decision. Sequencing tasks can be modelled by chaining a sequence of MLMs. MLMs have been used to generate clinical alerts and reminders, interpretations, diagnoses, screening for clinical research studies, quality assurance functions, and administrative support.

2.2.2 PROforma

PROforma is a formal executable logic language for describing clinical and other processes in terms of the decisions and other tasks that a physician needs to carry out to achieve its goals. It is capable of capturing the structure and content of a CPG in a form that can be interpreted by a computer. The language represents the basis for a method and a technology for developing and publishing executable CPGs. PROforma combines the features of formal specification languages as known in software engineering with the features of knowledge representation languages as known with artificial intelligence.

The PROforma language structure is based on a simple but versatile clinical process model known as the *domino model* shown in "Figure 1". This model derives from a variety of empirical studies of clinical decision-making and the development of aids to support patient management.

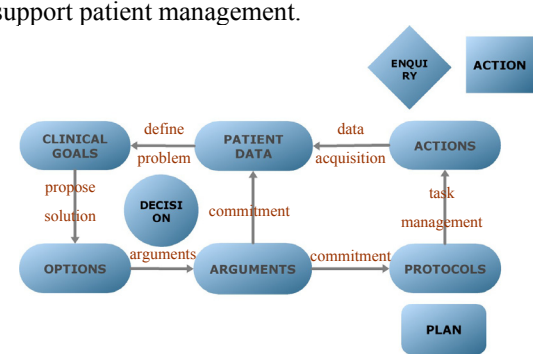


Figure 1: The relationship between the domino model of clinical process and PROforma tasks.

The left side of the diagram represents decision-making and the right the plan enactment. Given a set of beliefs, an agent may set certain goals and various solutions to these goals. With multiple options, such as alternative diagnoses or treatments, the agent must consider the arguments for and against these alternatives and make decisions based on the validity of each of the arguments. The domino model describes a relationship between actions, decisions, beliefs, plans, goals and candidate solutions, and the inference and processes linking them.

A further result was a reconstruction of the domino model into a minimal set of executable generic tasks: enquiries, decisions, plans and actions as shown in "Figure 1". Tasks are formal software objects that can be composed into networks representing CPGs or other processes, and it is from these tasks and the logical construct associated with each task, that the PROforma language is derived.

2.2.3 GLIF

GLIF is a computer-interpretable language for modelling and executing clinical practice guidelines. GLIF supports sharing of computer-interpretable clinical guidelines across different medical institutions and system platforms. It has a formal representation and defines the ontology for representing guidelines, as well as a medical ontology for representing medical data and concepts.

GLIF2 enables guideline modelling as a flowchart of structured steps, representing clinical actions and decisions. GLIF's guideline class also specifies maintenance information, the intention of the guideline, eligibility criteria, and didactics. The GLIF guideline instance syntax, which was based on a separately developed language, specifies the format of text files which contain GLIF-encoded guidelines. These files are used for sharing and interchange. However, the attributes of structured constructs are defined as text strings that can not be parsed, and such guidelines can not be used for computer-based execution that requires automatic collaboration.

In the year 2000 a new version of GLIF was introduced (GLIF3), an evolving revision of GLIF that attempted to overcome several of GLIF2's limitations. The GLIF3 model is object-oriented. It consists of classes, their attributes and the relationships among the classes which are necessary to model clinical guidelines. The model is described using Unified Modelling Language (UML) class diagrams.

2.2.4 Approach to Guideline Modelling

We were trying to examine the increasing sets of resources to obtain sufficiently amount of knowledge that is needed to design development tools and technique for building healthcare application that comply with the highest possible quality, safety and ethical standards.

To build an effective tool for capturing medical knowledge in a systematic and executable way among other criteria the following should be satisfied: access to the patient's record, use of standard medical vocabulary, clear semantics, knowledge maintenance and sharing, sufficient expressiveness to explicitly capture clinical processes of the guideline's author and leave flexibility during application runtime.

The primary goal of this project was to design an open source decision support application for decision-making between health professionals and patients.

2.3 Guideline Evaluation and Design

We have designed a machine interpretable guideline model as a sample of application that can improve inappropriate variations in practice, health care quality and to help control the costs. The guideline model was preceded by the profound study of the above mentioned guideline modelling methods. It includes an analysis of the literature and three published research projects concerning computerized specific representation models. We have tried to overcome the problematic of representing clinical knowledge in computerised manner by modelling clinical knowledge with CPGs. In achieving our goals we have followed the Stanford team dimensions: organization of guideline plan components, goals/ intentions, model of guideline actions, decision model, expression/ criterion language, data interpretation, medical concept model, patient information model.

These dimensions capture the essence of modelling the logic of computer-interpretable guidelines. The first four represent the core guideline components, and the last four link the guideline model to the patient data.

For identifying key elements of our CPG model the following elements were chosen:

Action specifies clinical actions that are to be performed during the patient's care process. These clinical actions can include diagnostic or therapeutic procedures that need to be performed in a treatment process, if an action step is triggered. Actions were modelled merely in a descriptive manner.

Enquiry describes the patient's temporary physical state. It contains the list of the attributes which are necessary for evaluating the patient at the temporary point of treatment. It looks into a patient's medical record and retrieves the values of the attributes described in the enquiry list, or asks the user for the required data.

Decision controls the flow of a guideline. It contains a group of candidates and a group of arguments for an individual candidate. Arguments are described by given weight. It supports inference in propositional and predicate logics, together with certain non-classic logic for reasoning and control of the guideline flow.

Plan components represent actions, decisions, enquiries, or hierarchical decomposed sub-plans of the guideline and their relationships. Plans merge together individual atomic tasks (decisions, actions and enquiries) or sub-plans into logical groups, control complexity and enable control of grouped elements.

Basic tasks that a guideline needs to satisfy are shown in "Figure 2". Data needed at the specific point of guideline execution is acquired by Enquiry. Based on defined values we can make decisions that evaluate patient's temporary physical state. Actions are defined for specific patient state, and are executed as needed.

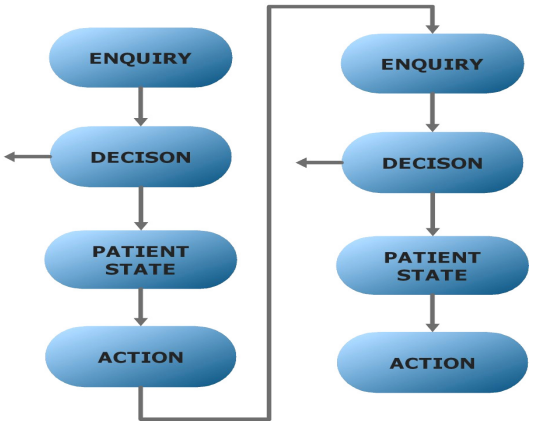


Figure 2: Logical flow of a guideline.

Constructs organization and decision model clearly defines goals of guideline and appropriate actions. The goal of a guideline flow is to choose an appropriate medical procedure for treatment of specific patient's state, by either excluding possible variations, or by choosing an optimal treatment process. Our decision support model supports the following features: each task can interrupt the guideline execution, some tasks are optional and can be ignored at execution time, some tasks need user interaction and have to be confirmed, all tasks can define preconditions that have to be fulfilled in order to execute that task, execution of each task brings system into a new state and sets new conditions, all tasks can have cycle execution, decision model can execute alternative tasks in parallel by excluding preconditions of each task, decision model is non-deterministic and uses arguments rules that can choose more than one candidate among available alternatives, expression language is represented using mathematical expressions which, when evaluated, give logical result *True* or *False*.

A guideline can be viewed in two abstract levels:

Conceptual level represents guideline as a flow chart, and is used for browsing and navigation through constructs. It represents a clear and simple overview of a guideline.

Executable level represents guidelines in full consistency and completeness. Definitions of enquiries items, clinical action and flow of the algorithm are specified at this level.

2.3.1 Database Support

Our model uses a support of relational database that is easy extendable and enables transformation or different representation of data. We have built an entity relational database diagram (ERD) with minimal set of entities as shown in "Figure 3". Starting-point of a guideline is the "*G*LGuideline" entity that describes general information about intention and its temporary state. For the moment we have four general tasks of guideline, the common attributes of which are collected in the "GLSteps" entity. Future extensions with additional type of tasks are easily deployable. Decision specific attributes are described with candidates and pertaining arguments. Enquiry specific attributes are described with predefined attributes and their pertaining values. All schedules that represent possible selection are modelled with separate entities.

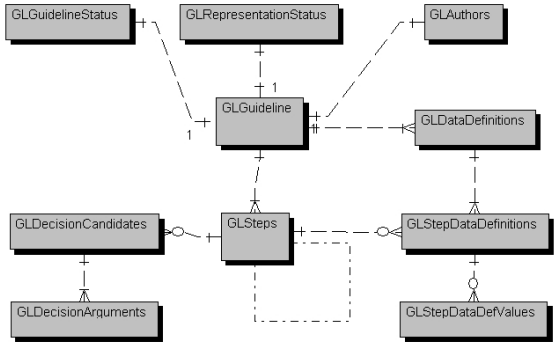


Figure 3: ERD diagram.

2.3.2 UML Design

We present a conceptual object oriented design of our application that capture the structure of guideline-based decision support tool in "Figure 4". The "DecisionProcessor" is a controller class that represents business logic and enables iteration through guideline steps. For evaluation of our expressions parameters and arguments are needed. "Candidate", "Argument" and "Parameter" are persistent classes. Postfix calculator can evaluate an expression with given arguments and parameters. The "DecisionStepState" enables the monitoring of each state of execution. Chosen actions and evaluated decisions are represented with classes "Action" and "Decision". In this sub-section a brief description of actual design is stated, so basically a more complex model was implemented.

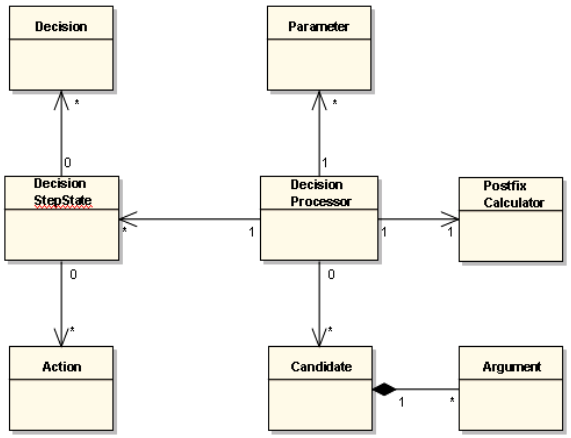


Figure 4: Conceptual class diagram.

2.3.3 Decision Model

The guideline modelling methods use a variety of decision models. The decision models ranges from simple if…then….else or switching constructs to complex models such as decision trees. Our model uses a decision tree, and selection among elements is modelled by mutually exclusive preconditions that depend on result of a decision task. The model uses argumentation rules for/against/choice for selection among alternatives. A none-deterministic choice is possible, where more than one alternative may be justifiable for a patient.

Expression language that defines preconditions, post conditions, cycles and candidates is defined as a complex expression containing operators, functions and variables. Operators are arranged into formula lexicon, term lexicon and unary lexicon. Each of these describes the operators available at certain parser priority. We have included 12 function lexicons which cover a wide range of standard functions, string functions, and other complex functions. Also, 2 operator lexicons are included. We have implemented more than 20 operators and more than 200 functions in an expression parser that is easily extendable with additional lexicon, or a new operator or function.

2.3.4 Constructing Guideline

We have implemented an information system based on CPG for modelling clinical knowledge. The result of this implementation is a graphical composer for creating machine interpretable guidelines. Those guidelines are fully implemented on executable level and can be executed to obtain appropriate clinical actions. The main window of the guideline editor is shown in "Figure 5". Top node of the tree represents

a main task that specifies information unrelated to the health knowledge, and is used for maintenance, change control, version control and other basic information of the entire guideline.

The enquiry element is used for acquisition of attributes that are needed at execution level of a guideline. Attributes are described as optional or mandatory, and their selection mode can be single or multiple. They can be predefined, and each of them ca be encapsulated within a specific attribute type.

Action task of a guideline is an important aspect, but in this project it has been modelled as simple textual description, and it is anticipated to be executed outside of the information system.

Decision uses a well defined decision processor with a rich mathematical expression language. Augmentation rules are used to choose one or more appropriate candidates. For each of them, we can specify a multiple arguments that, when evaluated, give logical result *True* or *False*. Each argument has a given weight. A selection control can be set to numeric mode or selective mode with the following rules:

- Rules that strictly exclude the alternative
- Rules that argue against the alternative
- Rules that argue for the alterative
- Rules that confirm or express strong preference for the alternative

A decision can also act as enquiry, in case it requires additional data for its execution.

Execution of tasks depends on their predefined precondition. Each executed task can inject a new condition into algorithm flow.

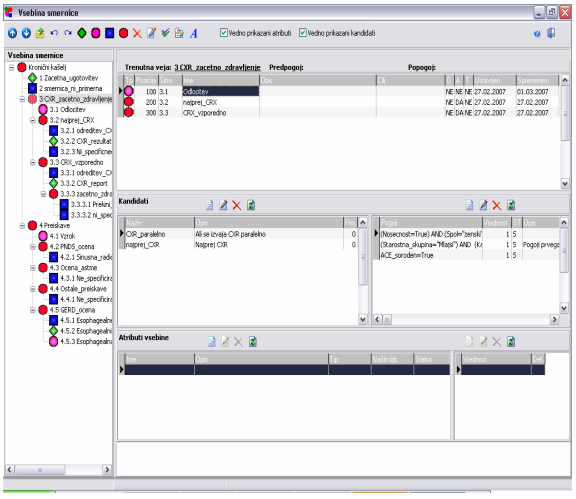


Figure 5: Guideline composer.

3 EXECUTING GUIDELINES

This subsection represents execution of guidelines modelled with our guideline composer. As a test of guideline execution, two guidelines were composed and executed for the purposes of testing. The first was treatment of chronic cough taken from the studies of PROforma, and the other one the determination and classification of hypertension disease taken from textual format using a book "Evidence based guidelines" and transformed into executable form.

At this point a guideline is represented at the executable level. The guideline execution is shown in "Figure 6". The left side of the picture shows a guideline content in execution process which provides the overview about which elements are chosen and which rejected. A user can also see a diary of events and check guideline for errors. During the runtime some occurrences need a user conformation which enables an interaction in guideline execution process. Also some tasks are optional and are to be confirmed or rejected. The left-top part of the "Figure 6" shows guideline attributes and their values which are valid at a moment. Also, chosen decisions and their evaluated candidates are shown for each execution step. In the right-bottom part appropriate actions are shown, together with their description and intention.

The intention of executing CPGs with this application is to obtain regularity of CPGs or to confirm already valid health care process.

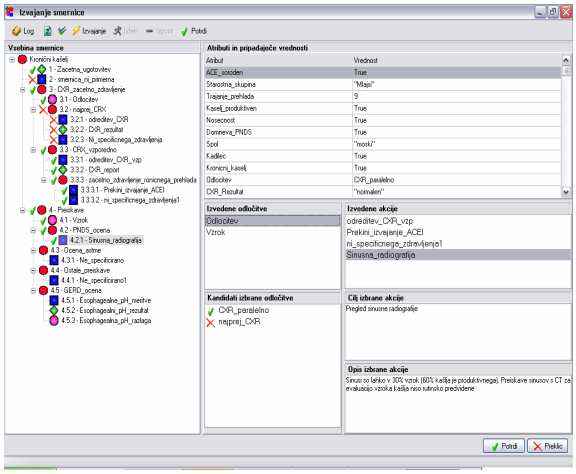


Figure 6: Guideline execution.

4 CONCLUSIONS

Growth of information about appropriate clinical treatment is enormous and makes its appropriate use in practice impossible. The purpose for building decision support systems for treatment process is to enable easy access to clinical knowledge. That is what should be the lead force for developing an appropriate and standardized CDSS.

By studying some of the many methods for representing clinical knowledge, guideline modelling and execution tools were developed. One of the most important aspects when developing a decision support tool is sharing information among other institutions which leads to a need to build a centralized data storage. For this purpose a relation database model has been developed and implemented. Model is flexible and fully extendable for further development. Application itself uses a multi user application server that enables sharing medical knowledge among users and institutions. For building a clinical guideline with composer a simple and easy understandable guideline constructs were implemented that are understandable to a person with none or poor computer knowledge.

Design of the application is object orientated and, if needed, extendable with other construct. Expression/criterion language uses a reach postfix mathematical parser. Among many already implemented operators and functions, it is possible to develop as many as needed user defined lexicons and inject them into the parser. This leads to a very flexible and adaptable expression language that can be used for complex decision making.

A newly proposed design represents an innovation in that it uses relational database support and a reach mathematical expression language parser which enables an infinitive and complex decision modelling.

For now, the application's primary goal is to build clinical practice guidelines and execute them in patient care process in order to obtain recommendable actions. Further development could lead to inductive learning, the statistical evaluation of effectiveness and appropriateness of guidelines by testing their regularity in a specific care procedure.

REFERENCES

Wyatt, J., Spiegelhalter, D. 1991. Field trials of medical decision-aids: potential problems and solutions, *Proceedings of the 15th Symposium on Computer Applications in Medical Care*.

Osheroff, J., Bankowitz, R., 1993. Physicians' use of computer software in answering clinical questions. *Bull Med Library Association*.

Hunt, D., Haynes, B., Segal, H., et al., 1998, Effects of computer-based clinical decision-support systems on physician performance and patient outcomes, JAMA, 280:1339–46.

Fox, J., Bury, J., Humber, M., Sutton, D., Integrating Bayesian inference into the PRO*forma* language, *Advanced Computation Laboratory, Imperial Cancer Research Fund, London*

Bury, J., Fox, J., Sutton, D., The PRO*forma* guideline specification language: progress and prospects.

Peleg, M., Boxwala, A.A., Ogunyemi, O., Zeng, Q., Tu, S.W., Lacson, R., Bernstam, E., Ash, N., Mork, P., Ohno-Machado, L., Shortliffe, E.H., Greenes, R.A., GLIF3: The Evolution of a Guideline Representation Format,1 *Stanford Medical Informatics, Stanford University School of Medicine, Decision Systems Group, Harvard Medical School, Brigham & Women's Hospital, Department of Medical Informatics, Columbia University*

Sordo, M., Ogunyemi, O., Boxwala, A.A., Greenes, A.R., Tu, S.W., GELLO: An Object-Oriented Query and Expression Language for Clinical Decision Support, *Decision Systems Group, Brigham & Women's Hospital, Harvard Medical School, Section of Medical Informatics, Stanford University School of Medicine, Stanford.*

Peleg, M., Boxwala, A., Modeling Clinical Guidelines in a Sharable and Computer-interpretable Way: Development, Implementation, and Use Requirements Instructors*, Stanford Medical Informatics, Stanford University School of Medicine, Decision Systems Group, Harvard Medical School, Brigham & Women's Hospital, Boston, MA.*

Tu, S.W., Campbell, J., Musen, A.M., The Structure of Guideline Recommendations: A Synthesis, *Stanford Medical Informatics, Stanford University School of Medicine, Stanford, University of Nebraska Medical Center, Nebraska Health Systems, Omaha*

Plavčak, A., 2007. Clinical Decision Support System, *Diploma work.*

NOVEL SENSOR TECHNOLOGY INTEGRATION FOR OUTCOME-BASED RISK ANALYSIS IN DIABETES

Mahesh Subramanian[1], Edward C. Conley[1,2], Omer F. Rana[1], Alex Hardisty[1], Ali Shaikh Ali[1]
Stephen Luzio[2], David R. Owens[2]
[1]The Welsh e-Science Centre, Cardiff University School of Computer Science
5 The Parade, Roath, Cardiff, UK, CF24 3AA and
[2]Diabetes Research Unit, Cardiff University School of Medicine
Academic Centre, Llandough Hospital, Penlan Road, Penarth, Cardiff CF64 2XX, UK
{m.subramanian, e.c.conley, o.f.rana, alex.hardisty, ali.shaikhali}@cs.cf.ac.uk; {luzio, owensdr}@cf.ac.uk

Steve Wright, Tim Donovan, Bharat Bedi, Dave Conway-Jones, David Vyvyan, Gillian Arnold
IBM United Kingdom Ltd, Hursley House, Hursley Park, Winchester, UK, SO21 2JN
{steve_wright, tim_donovan, bharat_bedi, conway, drvyvyan, arnoldg}@uk.ibm.com

Chris Creasey, Adrian Horgan, Tristram Cox
Smart Holograms, 291 Cambridge Science Park, Milton Road, Milton, Cambridge, UK, CB4 0WF
{chris.creasey, adrian.horgan, tristram.cox}@smartholograms.com

Rhys Waite
Zarlink Semiconductor Ltd, Phase 2 Mitel Business Park, Caldicot, Monmouthshire,UK, CF26 5YW
rhys.waite@zarlink.com

Keywords: Health informatics, home healthcare, biomedical sensor devices, mobility, wearable sensors, decision support system, individualised risk analysis.

Abstract: Novel sensor-based continuous biomedical monitoring technologies have a major role in chronic disease management for early detection and prevention of known adverse trends. In the future, a diversity of physiological, biochemical and mechanical sensing principles will be available through sensor device 'ecosystems'. In anticipation of these sensor-based ecosystems, we have developed *Healthcare@Home (HH)* - a research-phase generic intervention-outcome monitoring framework. *HH* incorporates a closed-loop intervention effect analysis engine to evaluate the relevance of measured (sensor) input variables to system-defined outcomes. *HH* offers real-world sensor type validation by evaluating the degree to which sensor-derived variables are relevant to the predicted outcome. This 'index of relevance' is essential where clinical decision support applications depend on sensor inputs. *HH* can help determine system-integrated cost-utility ratios of bespoke sensor families within defined applications – taking into account critical factors like device robustness / reliability / reproducibility, mobility / interoperability, authentication / security and scalability / usability. Through examples of hardware / software technologies incorporated in the *HH* end-to-end monitoring system, this paper discusses aspects of novel sensor technology integration for outcome-based risk analysis in diabetes.

1 INTRODUCTION

Continuous sensor-based monitoring technologies are central to new models of 'proactive' health and social care. In healthcare, 'proactive' implies a shift away from 'reactive' care – i.e. an 'illness-centric' model where interventions are made following presentation of symptoms or complications. The 'proactive' model embraces the World Health Organisation's (WHO) definition of health as "*a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity*" (WHO, 2007). To move towards this visionary goal, individuals need realistic

opportunities to improve and sustain their health and quality of life thus contributing to their own well-being. The availability of validated, relevant and ubiquitous personal healthcare information to minimise risk of predictable adverse events can empower and incentivise individuals to adopt more healthy lifestyles. Such technology can assist care throughout the 'patient path' (Abidi, 2001; Shnayder, 2005). Arising from these technology developments are significant ethical issues - e.g. in personal data protection and in establishing ethical authority for personal data reuse. We will discuss these issues in detail elsewhere in the context of our *Healthcare@Home (HH)* project - a research-phase generic intervention-outcome monitoring framework that integrates sensor-based technology as part of a disease early detection and prevention framework.

It is widely anticipated that future health information systems (HIS) will need to move from "institution-based" models to those that rely on 'near real time' data integration close to the patient. Interventions that use ethical risk stratification as part of a personal data integration framework is a priority in diabetes, where the number of affected individuals is predicted to rise from c.135 million people in 1995 to c.300 million in 2025 (King, 1998). All people with unmanaged diabetes are at substantially increased risk of serious medical complications such as retinopathy, kidney failure and peripheral neuropathy requiring limb amputation. As part of an individual's personal information management, the *HH* closed-loop model uses sensor-based trends to compute and stratify risk in a time frame and operational workflow that is meaningful and in a format that can be utilised for building decision support services (DSS). The DSS model in *HH* is founded on requirements of the Diabetes National Service Framework (NSF) standards for Wales and associated integrated care pathways (ICP). Section 2 summarises relevant related work. Section 3 describes technical aspects of the end-to-end *HH* system covering (1) smart sensors; (2) biometric authentication; (3) 'home hub' and (4) server-side architecture. Section 4 discusses concerns arising out of the project and possible future work with a conclusion in Section 5.

2 RELATED WORK

A healthcare technology platform utilising sensor devices can underpin comprehensive monitoring services outside of the hospital environment. This could support new ways of working that: (1) places less reliance on frequent clinical visits – subject to quality control / calibration safeguards and adequate clinical 'baseline' data (2) can incentivise patients to 'look after themselves' with realistic (achievable) personal guidelines within manageable episodes of care; (3) allows team-based caseload sharing between clinical visits to monitor progress and make escalation procedures robust (4) provides for the development of consistent risk prediction / longitudinal 'outcome recording' methodologies that are fit-for-purpose in scaleable *evidence-based* models (Williams, 2003; Conley, 2007). Several research projects address the issue of integrated care through the use of ubiquitous computing devices. S*APHIRE* (Hein 2006) is concerned with developing a healthcare monitoring and decision support system for cardiovascular disorders, assisted by wireless sensor devices in home settings. (Clemensen 2004) applied pervasive computing devices to the treatment / monitoring of diabetic foot ulcers. The *CODEBLUE* project (Lorincz, 2004) is typical of sensor device applications in medical emergency scenarios. In this project, micro-scale sensor devices (motes) (Crossbow, 2007) have been used to continuously monitor and wirelessly transmit vital sign data (e.g. heart rate and oxygen saturation data) to a data hub for processing. The CART (continuous automated real-time triage) system, developed by Advanced Health and Disaster Aid Network (AID-N) builds on the work carried out by the CODEBLUE team. A wearable tag has been developed which performs the following functions: triage, status display, vital signs monitoring, location tracking, information display and alarm signalling (Gao, 2006). The Smart and Aware Pervasive Healthcare Environment (SAPHE, 2007) supports telecare and lifestyle monitoring paradigms for early detection and prevention of adverse events i.e. for intervention before they become critical or life-threatening. Several remote healthcare monitoring systems are that currently use proprietary device information systems (e.g. Honeywell, American Telecare and AMD Telemedicine). Many more device families are expected to be developed in coming years conforming to global standards being established by the Continua Healthcare Alliance companies (Continua, 2007).

The *HH* system's conceptual driver is enabling 'near real time' risk analysis for early detection and prevention of disease. A Web Services-based platform to 'push' or 'pull' individual's health-related data along the patient path is being

configured in a manner that will reduce transcriptional errors. The end-to-end framework (Figure 1, see below) employs a collection of clinical hubs, mobile devices and / or dedicated home-based network servers to one or more data analysis engines.

3 END-TO-END FRAMEWORK

Figure 1 illustrates the conceptual design of an 'end-to-end' framework adopted in the *HH* project. The framework allows data capture from both wearable sensors and specialist hand-held instruments with wireless data transfer capability. Through a messaging fabric and / or dedicated integration application, raw data used in the risk analysis modules may originate from a wide variety of sources and device types (e.g. electronic forms, physiological monitors, retinopathy cameras, scanners, clinical chemistry or nucleic acid sequencing instrumentation). The data is integrated using a schema indicated as "QUID" (QUantitative Individualised Data integration) in the figure. In the context of an 'initial assessment' workflow, the types of diagnostic test to be performed are specified by a clinical registration procedure to be described elsewhere. The registration procedure has specific functions to ensure compliance to ethical protocols and informed consent procedures for re-use of data while verifying patient identity (see below). *HH* has adopted the Diabetes Continuing Care Reference (DCCR) dataset as the basis of its schema). Subsequent risk analysis of baseline data is undertaken in the context of a "disease model" - a quantitative research-based predictive framework that indicates *which* risk variables are most relevant to system-specified outcomes. In this regard, *HH* has been conceived as a comprehensive healthcare outcomes evidence-based *learning* platform.

3.1 Smart Sensors

In monitoring applications, the most common physiological analytes are blood gases (e.g. carbon dioxide, oxygen), blood electrolytes (e.g. potassium, sodium, and chloride), blood glucose, creatinine, urea, pH, cholesterol, bilirubin and proteins (e.g. albumin). The relative significance of values depends on the type of investigative scenario and progression status of disease. In diabetes, the measurement of glucose concentration at an appropriate frequency is routine in short-term prevention of hyper- and hypo-glycaemic events and in long-term prevention of complications. Automated measurements can be enabled to support continuous monitoring. Conventionally, glucose monitoring is performed using 'finger-stick' devices. These provide only intermittent measurements and rely on patient compliance. Next-generation glucose

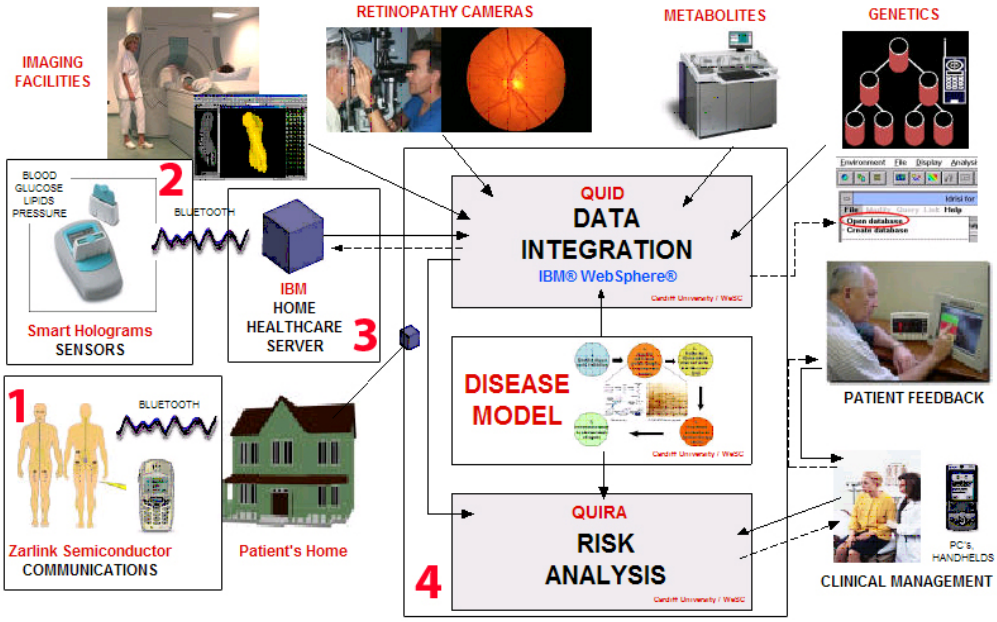


Figure 1: *HH* end-to-end framework.

monitoring devices aim to provide real-time continuous measurements by non- or minimally-invasive means. One such sensor designed for integration into the *HH* system uses glucose-responsive holograms based on thin-film polymers incorporating phenylboronic acid receptors. Selective binding of glucose to the receptors induces swelling or contraction of the film. This physical change in turn causes the spacing between holographic fringes within the film to increase or decrease, thereby modulating the colour of the light diffracted according to Bragg's Law.

To obtain repeatable quantitative measurements (figure 2), the colour of the hologram can be read with a portable optical reader equipped with Bluetooth®. These sensors exhibit long-term chemical and physical stability, enabling measurements over long time periods without evidence of hysteresis. A further advantage is the ability of holographic analyte sensors to be incorporated into a multitude of formats (e.g. catheters, contact lenses, implants), thereby offering increased patient choice.

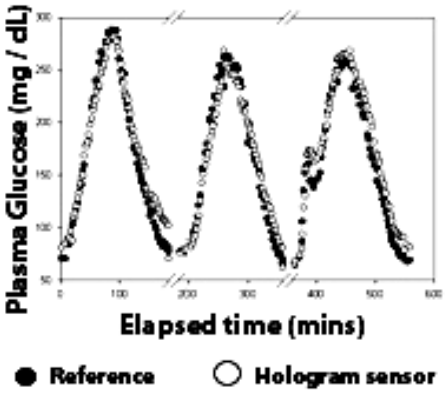


Figure 2: Real-time measurement of blood glucose using a holographic sensor compared to measurements made off-line and post experiment by a traditional reference method.

The prototype hand-held device (figure 3) has a touch-screen user interface. The reader employs a bespoke spectrometer optical sub-assembly, combined with analogue to digital converters (ADCs). These ADCs are read by a microprocessor which provides a calibrated output of wavelength and the corresponding concentration of the analyte being measured. This is achieved by use of look up tables or calibrated polynomial curve fit data. Temporal data trends can be calculated on-device or data can be sent via Bluetooth® to a *HH* web service to enable downstream decision support.

Figure 3: Smart Holograms prototype Hologram reader with integrated Bluetooth® connectivity.

3.2 Biometric Authentication

The *HH* system design has evaluated scenarios where people entering data might be easily confused by similar names or as different members of the same family. A number of protocols for disambiguation and absolute patient identification can be incorporated, including biometrics and / or smartcards. Biometric variables can be classified either as physiological (e.g., derived from a fingerprint, face or iris scan) or behavioural (e.g., speech recognition) (Biometrics, 2007).

The *HH* system requires technology options that are cost-effective, fast and accurate. Based on these criteria, biometric identification used robust fingerprint recognition technology ('2' in figure 1) (Wilson 2003). For flexibility, the biometrics device was designed to be "loosely-coupled" with the sensor and the home hub modules (refer to section 3.3). This permitted different packaging options with integration of new sensor types without substantial additional engineering cost. The hub is used to tag the biometric ID to the incoming sensor data in order to form an association between a patient identity and a data reading. Communication between sensor, home hub and the biometrics module is *via* Bluetooth®, using a protocol specified by IBM, Zarlink and Smart Holograms.

In its current operational mode, the biometric device (figure 4) used in the *HH* system saves the scanned fingerprint of a user in local memory, and the individual is prompted to assign an "ID" specific to that fingerprint. That ID is then transmitted *via*

Bluetooth® to the data hub to be tagged within an electronic patient record. This "enrolment" process can be used prior to sensor readings to validate identity. A delay-free smartcard that has high end-user acceptance (e.g. contact-less 'wave & pay' cards) can also be used in appropriate circumstances e.g. to enable inter-service access along the patient path. Identification of the patient by means of their enrolled ID sends that ID to the patient-proximal hub (in the clinical data 'baseline data' operational hub or the patient's own home hub, or mobile hub). The home hub can associate the subsequent sensor reading with the patient ID. This is the "identification" process. The design of the device ensures security of patient information. All scanned fingerprints are stored in the memory of the biometric device. No fingerprint scans are transmitted. All patient-identifiable information can be encrypted at source, in transmission and storage. If a hacker were to capture the wireless transmissions the data would have no meaning.

Figure 4: Fingerprint scanner by Zarlink.

3.3 Home Hub

The function of the 'home hub' ('3' in figure 1) is to collect and collate the data from sensor(s) and the biometric and smartcard / reader device(s) and to co-transmit these *via* an appropriate communication channel to the remotely-located server. The home hub sits at the centre of the data collection and transmission capability of the system, where all devices - sensors, authentication module and server connect through common interfaces. Various physical realizations of the hub are possible. It can, for example, be a mobile device such as a standard mobile phone (we have demonstrated functionality on a Sony Ericsson P910 phone - figure 5) that can provide near real-time data connectivity. The hub can also be deployed as a fixed 'wired' hub using, for example, Ethernet connectivity. In a clinical environment this also provides near real-time data connectivity. For home use, hubs can be configured to upload data periodically e.g. once or twice a day.

Figure 5: Mobile phone personal data hub by IBM.

Sensor devices currently connect to the hub through adapters specific to the type of interface required. Adapters are device transmission protocol specific and can be developed by any manufacturer wanting to provide connectivity of their devices to this infrastructure. In *HH*, the medical devices use Bluetooth®, although the architecture allows for this to be any available. Zigbee has some power consumption advantages over Bluetooth® (Zigbee, 2007).

We have developed adapters for a variety of devices that can be used in conjunction with the demonstrator system; namely for: (1) weighing scales; (2) blood pressure cuff; (3) pulse oximeter; (4) glucose meter. In addition to adapters for sensor devices, an adapter is also required to interface the hub to an appropriate application server. This step is also protocol-specific and in practice a range of adapters may be needed according to specific application scenarios.

The hub architecture (figure 6) consists of an event engine that behaves much like a broker, in that it can receive events from one adapter and passes these onto another adapter. In the simplest operation, the event engine receives events from a single sensor device (for example, weighing scales) and sends this event to the adapter that transmits these events to an application server. The current implementation of this architecture uses the IBM Personal Care Connect toolkit (Blount, 2007) and is based around a number of standard technologies. These include:

- Java 2, Micro Edition, Mobile Information Device Profile, Connection-Limited Device Configuration (J2ME, MIDP, CLDC) or OSGI for fixed hub (Java ME, 2007).

- JSR-82 to allow Java to interface to Bluetooth® (assuming Bluetooth-enabled medical devices) (JSR 82, 2007).
- Bluetooth® or mobile connectivity via GSM / GPRS for wireless hubs.
- Ethernet, ISDN, etc. for fixed / wired hubs.

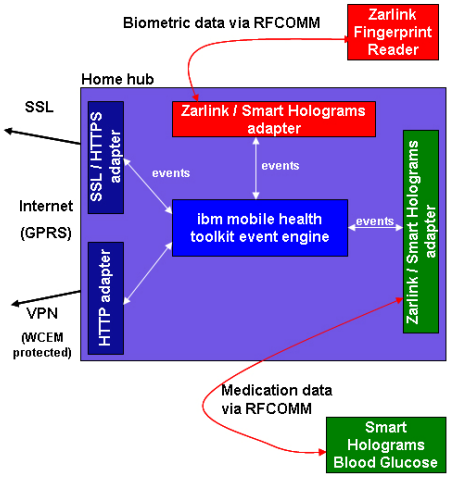


Figure 6: Hub technical overview.

The hub design supports multiple adapters – key to creation of an open-hub platform. Each adapter is specific to a *hub family*, which is characterised by the following attributes: (i) device-to-hub communication, (ii) event data representation, (iii) event serialisation, and (iv) hub-to-server interaction. Different hub families can use different event classes, communication protocols, and serialisation schemes. A hub family may be optimised to handle a specific set of biomedical sensor types. Each adapter handles communication with a family of hubs by (i) defining a protocol, (ii) de-serialising the data forwarded by the hub, and (iii) extracting event data from the event after instantiating the events. Adapters also convert proprietary data events into a consistent format – i.e. act as a definable interface. It is expected that *HH* will migrate to industry recommendations of consortia such as the Continua Alliance (Continua, 2007). The hub functionality is currently implemented in Java and is deployable to a mobile phone as a MID-let or to a fixed hub as a set of OSGI bundles.

In the current research demonstrator, sensor data can be obtained from devices and stored on the hub prior to transfer by a standard HTTP protocol. A HTTP 'Post' agent uses stable storage to cache events that cannot be sent immediately and then forwards them when the next connection to the server is established. Data sent to the server is secured using the Secure Sockets Layer (now referred to as Transport Layer Security) (SSL, TLS). Other options include: Access Point Node (APN) and Virtual Private Network (VPN).

Secure Sockets Layer: SSL is commonly used within internet applications to provide secure client to web server connections.

Access Point Node: The Access Point Node (or Name) is the definition of the internet connection on a GPRS mobile device that provides *the route the data will take* from the GPRS device to access other networks such as the Internet. All APN's are defined within the mobile device with a username and password. The APN concept assumes the existence of GPRS support nodes that exist between a mobile device (using a GSM or UMTS service) and a server providing IP-based access to the Internet. By obtaining a private APN (with its own unique credentials) it is possible to create a private connection back to the *HH* servers. The *HH* demonstrator has been successfully tested with APN.

Virtual Private Network: A VPN effectively creates a private network by creating a secure 'tunnel' through the existing IP network. The *HH* demonstrator has successfully used the IBM VPN product *WebSphere Everyplace Connection Manager (WECM)*.

A sub-class of specialised functions that mimic adapters exist to create functions internal to the hub. A number of these have been implemented within the current demonstrator. For example, an "Audio Alert Agent" makes pre-configured audio alerts in response to certain events like the reception of data from a device or successful transmission of an event to the server. The "User Display Agent" enables the display of status information and provides input capability for the user.

3.4 Server-side Architecture

The server-side architecture comprises of a 'QUID' and 'QUIRA' component (figure 7 and '4' in figure 1) and has been described in detail elsewhere (Subramanian, 2006, Shaikh-Ali, 2007, Conley, 2007).

The QUID component focuses on delivering data collection, data storage, process execution and portal

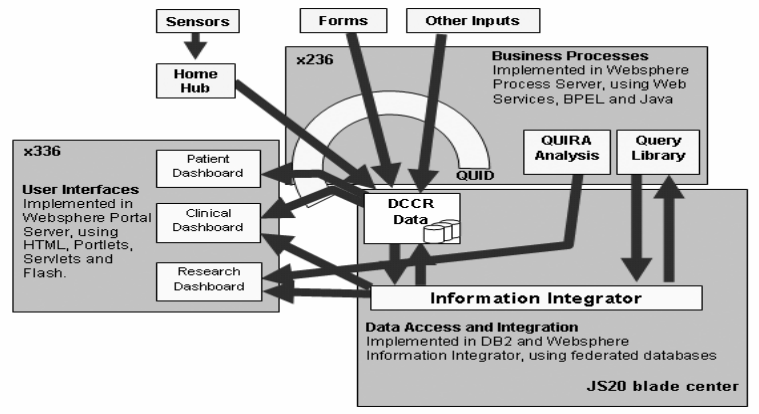


Figure 7: Physical implementation of the conceptual design shown in Figure1.

infrastructures to support users of the system including clinicians, patients and researcher roles. The development of QUID has been guided by the requirements of Integrated Care Pathways (ICPs) for diabetes (ICP, 2007). The QUID component collects, validates and stores data streams from hubs along the patient path. The interface for data presentation / collection emulates workflows in real-world care pathways. This 'end-user familiarity' design feature is strengthened by presentation of the sensor data in a timeline-based (longitudinal) layered manner, make the data meaningful. We have used OpenLaszlo (Laszlo, 2007) to display charts in an intuitive manner. Similarly, the patient portal permits access to personal data supporting the self-management paradigm

The QUIRA (Quantitative Individualised Risk Analysis) component comprises a risk analysis engine performing various operations identifying signals in the longitudinal data stream, alerting the care team to fulfilment of pre-defined risk criteria. QUIRA represents ongoing research we will report on in the future.

4 CONCERNS & FUTURE WORK

"Pervasive" or "ubiquitous" computing covers a range of research topics, including distributed computing, mobile computing, sensor networks, communications, artificial intelligence, and human-computer interaction. It is an emerging field of research, and as such has many unresolved issues – notably in areas like security, usability, privacy and ethics amongst others. In the framework described in this paper, ubiquitous computing devices demand secure transmission of data to the server, in turn demanding encryption mechanisms that defeat purposeful or accidental 'eavesdropping'. Device miniaturisation assists the resolution of powering issues. Significant ethical and privacy issues remain, and it is axiomatic that a comprehensive informed consenting process needs to be developed that is fit-for-the defined purpose of disease early detection and prevention. Informed consent needs to be properly structured in the recording workflow.

There is currently no widely-accepted standard protocol for the device-to-server data transmission and / or format / structure for data being transferred between devices. We expect interoperability between devices from different vendors to be a key focus within the scope of the Continua Healthcare Alliance (Continua, 2007). Irrespective of global technical standards adopted, methodological standardisation of data acquisition needs to be defined in order for patients and carers to reap the benefits of interoperable systems.

The *HH* project has to date employed non-invasive sensors. However we anticipate significant developments in coming years in the area of invasive (implantable) sensors. Implantable sensors will likely be micro-miniaturised devices that can be implanted into a patient's body to enable relaying of health-critical signals on a semi-continuous basis.

5 CONCLUSIONS

System-based management of chronic conditions is essential to improve healthcare outcomes. Conventional models of healthcare provision lack capacity to continuously monitor physiological data

combined with life event ‘timelining’. In such an information system, healthcare can be provided in a ‘patient-centric’ model that maximises healthcare resources.

ACKNOWLEDGEMENTS

This project has been funded by the Inter Enterprise Computing Theme of the UK Department of Trade and Industry (DTI)-led Technology Programme, for which we are grateful. ECC and DRO are grateful for support by the Wales Office of Research and Development for Health and Social Care, Wales Assembly Government.

REFERENCES

Abidi, S.S.R, 2001, “An Intelligent Tele-Healthcare Environment Offering Person-Centric and Wellness-Maintenance Services”, Journal of Medical Systems, Vol. 25, No.3.

Biometrics, a, date visited: July 2007, link visited: http://en.wikipedia.org/wiki/Biometrics

Blount, M., Batra, V.M., Capella, A.N., Ebling, M.R., Jerome, W.F., Martin, S. M., Nidd, M., and Wright, S.P., 2007, “Remote health-care monitoring using Personal Care Connect”, Volume 46, Information-Based Medicine, IBM Systems Journal.

Conley, E.C., Owens, D.R. Luzio, S., Subramanian, M., Shaikh Ali, S., Hardisty, A., Rana, O. (2007) “Simultaneous Trend Analysis for Evaluating Outcomes in Patient-Centred Health Monitoring Services”, Springer Journal of Healthcare Management Science, Submitted.

Clemensen, J., Larsen, S. B., and Bardram, J., 2004, “Developing Pervasive e-Health for Moving Experts from Hospital to Home”, Proceedings of the IADIS e-Society Conference, pp.441-448, Avilla, Spain.

Common User Interface, June 2007, link visited: https://www.cui.nhs.uk/Pages/NHSCommonUserInterface.aspx

Continua Healthcare Alliance, June 2007, link visited: www.continuaalliance.org

Crossbow, June 2007, link visited: www.xbow.com

FIPS, 2002. “Security Requirements for Cryptographic Modules”, link accessed http://csrc.nist.gov/cryptval/140-2.htm

Gao, T., Massey, T., Sharp, J., Bishop, W., Bernstein, D., and Alm, A.., September 2006 “Integration of Triage and Biomedical Devices for Continuous, Real-Time, Automated Patient Monitoring”. Proceedings of the IEEE Medical Devices and Biosensors Conference (IEEE MDBS 2006), Boston, MA..

Hein, A., Nee, O., Willemsen, D., Scheffold, T., Dogac, A., Laleci, G., 2006, “SAPHIRE - Intelligent Healthcare Monitoring based on Semantic Interoperability Platform - The Homecare Scenario”, 1st European Conference on eHealth (ECEH’06), Fribourg, Switzerland.

Java ME: Mobile Information Device Profile (MIDP), date visited: July 2007, link visited: http://java.sun.com/products/midp/

JSR 82: Java APIs for Bluetooth, link visited: http://www.jcp.org/en/jsr/detail?id=82

King, H., Aubert, R., Herman, W., 1998, “Global burden of diabetes, 1995-2025: prevalence, numerical estimates, and projections” Journal: Diabetes Care,Volume: 21 Issue: 9, pp 1414-31.

Lorincz, K., Malan, M., Fulford-Jones, T., Nawoj, A., Clavel, A., Shnayder, V., Mainland, , S., Welsh, M., Oct-Dec 2004, “Sensor Networks for Emergency Response: Challenges and Opportunities”, In IEEE Pervasive Computing, Special Issue on Pervasive Computing for First Response.

Laszlo, date visited: July 2007, link visited: http://www.laszlosystems.com/.

ICP, Diabetes NSF (Wales) Standards, link accessed: www.wales.nhs.uk/sites3/home.cfm?orgid=440

Shnayder, V., Chen, B., Lorincz, K., Jones, T., & Welsh, M., 2005, “Sensor Networks for Medical Care”, Technical Report TR-08-05, Division of Engineering and Applied Sciences, Harvard University.

Shaikh-Ali, A., Rana, O.F. Hardisty, A., Subramanian, M., Luzio, S., Owens, D.R. and Conley, E.C., October 2007, "Portal Technologies for Patient-centred Integrated Care", European Conference on eHealth, Oldenburg, Germany.

SAPHE project, July 2007, link visited: http://ubimon.doc.ic.ac.uk/saphe/m338.html,

Subramanian, M., Shaikh-Ali, A., Rana, O.F. Hardisty, A. and Conley, E.C., October 2006, "Healthcare@Home: Research Models for atient-Centred Healthcare Services", John Vincent Atanasoff International Symposium on Modern Computing, Sofia, Bulgaria. IEEE Computer Society Press.

WHO, World Health Organisation, date visited: July 2007, link visited: http://www.who.int/en/

Wilson, C., Hicklin, R.A., Korves, H., Ulery, B., Zoepfl, M., Bone, M., Grother, P., Micheals, R., Otto, S., Watson, C., 2003, “Fingerprint Vendor Technology Evaluation”, National Institute of Standards and Technology, Mitretek Systems, NAVSEA Crane Division

Zigbee, date visited: July 2007, link visited: http://www.zigbee.org/en/index.asp

A NEUROCOGNITIVE PROTOCOL SYSTEM TO SUPPORT HEALTH AND CARE OF ABUSED CHILDREN

Carlo Emmanoel T. de Oliveira, Carla Verônica M. Marques
Programa de Pós Graduação em Informática, Universidade Federal do Rio de Janeiro
Av. Athos da Silveira Ramos, 274, Rio de Janeiro, Brasil
carlo@nce.ufrj.br, carlamarques@posgrad.nce.ufrj.br

Jorge Neval Moll Neto
Unidade de Neurociência Cognitiva e Comportamental, Rede LABS-D'Or Hospitais, Rio de Janeiro, Brasil
mollj@neuroscience-rio.org

Keywords: Neuropsychology, neurobiology, abused children, assessment, datamining, decision support systems, web development.

Abstract: Abused children is highly endangered of developing critical cognitive dysfunctions. Clinical observation has encountered many related cases of abuse and poor learning performance. Authorities unaware of these conditions may take longer to act, detrimentally to the child welfare. This work provides a wide coverage of medical protocols for every area concerned with endangered children procedures. These protocols were researched with the collaboration of specialist in each area to achieve the most detailed and conspicuous information of children status. These protocols are proposed as a Web system available to all concerning professionals and authorities to input and access the relevant information. This data can be processed and analyzed to provide decision support and handling indications derived from statistical and heuristic treatment of the whole information.

1 INTRODUCTION

The violence in the childhood is highly co-related with serious behavioral, cognitive or emotional damages that are immediately noticed on school learning, on language expression and on the relationship life. The abuse and negligence commit the social cognition development, that means to say that the semantic and pragmatic dimensions of the language become very impaired, what leads to the delay in the vocabulary acquisition and development and in grammatical structures of the oral and writing language.

These results found in the neuropsychological tests, show qualitative alterations, mainly in tasks that involve the activity of the frontal lobe. On first analysis, poor performance was observed in sheltered children, in levels, co-varying the type, intensity, duration of the abuse and/or negligence, age group in which elapsed the abuse situation and presence of aggravating environmental factors (shelter type or street experience) or the opposite, opportunities of compensatory interpersonal entails. Qualitative differences were identified on the cognitive development, especially in the attention, formal learning, memory, language, abstract reasoning and executive functions, without deep lowering of the global cognitive competence (IQ).

In the same way, the behavioral alterations in sheltered children, identified as risk factor for the development of psychiatric impairment, appeared to be related to the abuse conditions, abandonment or negligence. This symptoms varied a lot depending on the anxiety, shyness, phobias, panic syndrome, impulsiveness, low-self-esteem, little or no tolerance to the frustration, disturbed sleep, night enuresis and presence of psychogenic motor stereotypies.

Behavior disorders - post-traumatic stress syndrome, deficit of attention disorder, psychotic symptoms, obsessive-compulsive disorder and conduct disturbances, little interpersonal ability, acceptance anxiety, imitative behavior, and even dissociatives disorders: hiperactivity, hipervigilance, threats illusory perception, paranoiac interpretation of interpersonal relationships, emotional immobility, incapacity of reacting when challenged or in

pressure, inferiority and uselessness feeling, suicidal and/or homicidal ideation, hallucinations, irritation, despair and self-mutilation.

And finally dysthymia: chronic depressive state, melancholy, obsessive-compulsive behavior, bad-humor, low motivation, low self-esteem, emotional apathy, pessimism, anxiety, chronic fatigues, self-isolation, alimentary and drug addiction disorder.

We still have academic difficulties in the learning of school contents in the Portuguese language and mathematics areas.

Our investigations show that sheltered, abuse or negligence victims, children have low performance at school, revealing that the consequences of violence can persist years after their retreat from the streets or from the origin family. According to the collected information, it happens because the child's past traumatic experiences seem to be added to the sheltering situation in which they are. Besides to total absence of specific educational programs for these children, they make their formal learning impossible, becoming an apart group inside the school. The life conditions in the shelters become an aggravating factor on symptoms of cognitive delay and school abandon, appearing as the cause of high stress and propitiating the continuity of the privation conditions and abuse, that end up resulting in the escape of children from shelters and in the consequent school escape. As well as the school reaffirms these differences becoming the official organ that decrees the failure and the social exclusion.

2 OBJECTIVE –SUPPORTING ENDANGERED CHILDREN

TUIA is a computational program constituted of a base of organized data to recover and to co-relate information associated to the abuse, abandonment and negligence experiences against the child or adolescent. It was projected to facilitate the interaction among specialists of different areas interested in child abuse, and also encouraging researchers to collect information from multiple services. The compilation of these intersubject information allows to cross check information from the various data sources, reinforcing more subtle indications. Cross checking medical and neurocognitive data can exemplify the case. Data measuring child abuse experience from medical sources can be co-related the neurobiological performance (De Bellis MD, 2005), neuropsychological (Beers SR, De Bellis, 2002), educational and the child's physics. The combination between the two modules aims to find the sequels indexes, esteeming the relative direct consequences to the abuse, abandonment and negligence indexes.

The program enhances the handling endangered children, allowing the concomitant examination of multiple aspects. Traversal of the whole information database can unravel hidden indications, which are not directly accessible, but can be only inferred through the intelligent comparison of the parts. An information system integrated in a national ambit, should be projected to be used jointly with the specialists and authorities, promoting an appropriate sequence of the institutional actions for prevention, identification, evaluation and attendance of this population (Stemberg,K.J.,2004).

3 REVIEW – THEORETICALS

There is now abundant evidence that childhood abuse and neglect can result in permanent changes to the developing human brain. These changes in brain structure and function appear to cause psychological and emotional abnormalities during childhood and adulthood. Behavioral and psychological problems include impulsive and instrumental aggression, learning disabilities, mood disorders, post-traumatic stress disorder and antisocial personality disorder, among others.

Modern neuroimaging techniques, among which magnetic resonance imaging (MRI) methods stand out as the most promising ones. MRI not only is non-invasive (no ionizing radiation or other biological effects have been shown), but it provides powerful ways to directly address brain structure and function in detail both in transversal and in longitudical studies.

Voxel-based morphometry, for example, is an especially interesting technique which allows the study the structure of virtually the whole brain in a statistically robust manner. By comparing control and patient groups paired by demographic variables (e.g., age, education, physical developmental measures), statistical maps of grey or white matter changes can be generated (Good et al., 2001). Furthermore, continuous variables can be entered as parameters, allowing for testing not only categorical differences among groups, but also dimensionally. Voxel-based-morphometry has been successfully employed to detect subtle anatomical changes in neurologic and psychiatric disorders, such as major depression (increased amygdala volume), antisocial personality disorder (reduced anterior temporal lobe

volume), among others (for a review, see Meyer-Lindenberg and Zink, 2007). Additionally, a new MRI technique dubbed diffusion tensor imaging has been developed during the past 10 years. This modality allows the determination of microscopic water molecule flow (Brownian movement) in the living brain tissue; further developments have used mathematical models of water diffusibility, and now enable researchers to trace the 'brain wiring', or the white matter fibers, based on inferences from preferential water diffusibility. This technique allows for direct visualization of developmental changes in the human brain secondary to genetic and environmental factors. Results from our lab showed that neuroplastic changes can readily be detected using this method (Tovar-Moll et al., 2006).

Finally, functional MRI is another especially powerful non-invasive technique with demonstrated robustness in detecting functional reorganization of the human brain following developmental and cognitive-emotional factors. Our group has been using functional MRI now for about 10 years in the study of complex cognitive-emotional phenomena, which include moral judgments and moral sentiments (Moll et al., 2005). Based on lesion evidence and functional MRI experiments in adults, we have formulated a now influential model of the 'moral brain'. Functional MRI and available cognitive models can now provide important guidelines for studies addressing cognitive-emotional development in normally developing and abused or neglected children. This will certainly be a fruitful line of investigation for our understanding of the complex cognitive and emotional disturbances in these children, which can guide the development of better assessment and treatment schedules aiming to prevent further brain damage or ameliorate established symptoms. Finally, these imaging techniques can be employed together with genetic studies (e.g., gene polymorphisms, endophenotypes, etc), providing an unique opportunity to explore not only structural-behavioral or genetic-behavioral interactions, but direct genetic-structural effects (ex., Meyer-Lindenberg and Zink, 2007).

4 ORGANIZATION – A COLLECTION OF PROTOCOLS

Thirty-two children were appraised in the age group between 7 to 12 years, residents of Ayrton Sena shelter, of the city hall of Rio de Janeiro, that possess official (juridical, technical or administrative) registration of abuse, negligence or abandonment history and 32 children of control group who go to the same schools the sheltered children go and who live with their biological families in a common home. There is no suspect or record, even informally, of abuse or negligence, by relatives and teachers. This group is similar in age, sex, socio-economic level and school levelling.

The sheltered children had a battery of investigation instruments different from the non-sheltered children referring to the life history, family context data and to the current psych-social situation. It was necessary to diversify, using an abuse checklist (Joseph Pitty) and a specific medical history assessment, in the first group, and an inventory of refined qualitative analysis of family relationships associated to a general medical history assessment, appropriate for the second group. Such procedure looked for comparing different instruments, similar in objectives and adjusting them to measure similar information in importance, relative to its different contexts. In this phase, socio-economic questionnaires and an environmental analysis questionnaire were also applied to evaluate the shelter, and the conventional pediatric and neurological exam.

The exam of psychiatric impairments tracking, the CBCL - (ASEBA) - Child Behavior Checklist (CBCL), the Parent Report Form, and the Teacher Report Form are part of the first evaluation stage. In the positive cases of this screening, some information were explored based on chosen question from K-SADS-PL - diagnoses interview for children and adolescent between the ages of 6 and 18 – Brazilian version of Schedule for Affective Disorders Schizophrenia for School Aged – Children.

Based on these general data, the children with mental and sensorial deficiency were excluded. The children with complications at birth, serious diseases, lesions and wounds in the head, internments in ITC, history of comatose state, previous evaluation of IQ with index below 80, history of treatments with psychotropics, psychiatric impairment, alcohol or drugs abuse, dependence or prenatal exhibition to alcohol or substances, were not excluded, but considered as fundamental part of the research by dealing with most of the subjects.

After this stage, a pedagogic evaluation (CESGRANRIO) of Portuguese and mathematical language was applied jointly with the complete neuropsychological battery, including the language areas (Capovilla battery), moral competence (Moral

Judgment Test (MJT) - Georg Lind.) and cognition using Wechsler Intelligence Scale for Children (WISC III), Tower of London (TOL), Children's Color Trail Test (CCTT), Test of Cerebral Dominance (BTN) and Span Cores (short term memory).

The language evaluation was applied in three sessions of specific abilities exams composed by the vocabulary Test in images, Test of phonological discrimination, Test of words and pseudo-words repetition, Dictation, nomination Test, Test of letters knowledge, Test of fluency of words, Proof of phonological conscience, Proof of syntactic conscience, Test of silent reading competence and by the Test of understanding of written sentences.

The analysis of these preliminary data served as foundation for the development of a protocol of structural (morphological) and functional investigations of the brain using the magnetic resonance imaging (MRI) techniques.

5 IMPLEMENTATION – A WEB SYSTEM FOR CHILDREN CARE

An effective way of caring for endangered children is tracking closely all developments in theirs lives and handling readily this information to the authorities responsible for their care. Many children continue exposed to critical situations due the lack of perception of health and education authorities of their particular problem. This work propose to mitigate this risk by gathering comprehensive information on critical aspects of children conditions and making this information accessible within a decision support tool.

The application is a Web based system with lightweight componentization and a flexible protocol reconfiguration architecture. The main purpose is to support the creation and maintenance of clinical protocols that can gather the most conspicuous information about endangered children. The protocols are classified into team packages and into specialist forms inside each team division. Team packages can be assemble to concentrate the practices requiring a specific expertise. The packages contain items covering the whole area of the team expertise and consist mainly of engineered information collectors that can be readily processed into decision support reports. In the use cases shown below, researchers can devise protocols and analysing algorithms to provide the state of art children care environment. Institutional teams can apply the new protocols and analyse the reports to allocate children to the proper treatment.

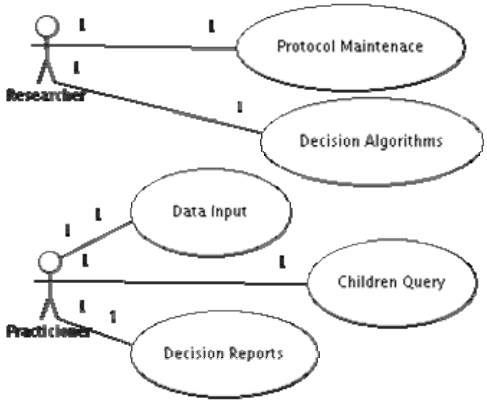


Figure 1: Main Use Cases.

The tool not only supports mainstream handling of current children data, but also is ready to incorporate new research to the immediate benefit of these children. The tool includes an upgrade mechanism to upload new protocols and decision support algorithms to handle new information and requirements to cover current children handling needs. The system is designed to keep up with the evolution of people caring knowledge, supporting protocol modifications while preserving previously collected information. The domain model was designed taking up a thee years study, covering the various involved areas and specialists which contribution was analyzed to develop a common denominator. The simplified diagram below shows the tool model, supporting the creation of new protocols capable of gathering children data. Reporting tools can also be attached to the system to provide decision making graphs about a resulting query.

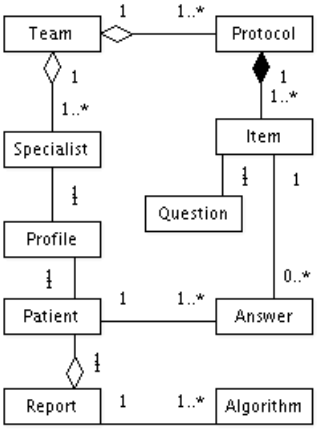


Figure 2: Static Domain Diagram.

The system can be hosted in more than one place and the packages and its information can be componentized through JSON (JSON, 2007) and REST (Khare & Taylor, 2004) communication protocols. Information is transacted across sites using REST requests devised as a DOM (DOM, 2007) like interface. Since information about a patient can be entered in the various hosts maintained by the specialist institutes, each host can request complimentary information to the local specialist data. The requested information is transacted through JSON packets in response to the REST requests.

The application has been given the name TUIA (Test Unification for Indications of Abuse) and is built upon Web 2.0 techniques, to enable simple construction, social interaction, and componentized structure. The idea is to provide easy integration with other systems composing the TUIA system in special the neurocognitive learning acceleration objects. These objects are neuropedagogical games designed to reduce the gaps between strong and weak cognition functions.

The interdiciplinary data is stored in the system database and is processed by algorithms in the machine learning engine. This engine provides inference about the healthcare subjects and classify information to help medical, judicial and political decisions.

The current engine installed is a Bayesian clustering (Binder, 1978) algorithm using the Orange framework. As a proof of concept, a initial set of 44 children was submitted to a clustering analysis. This groups was submitted to a battery of language, mathematics and attention tests. The group is originally composed of 29 children under state custody and a control group of 15 regular school children. The initial groups were assigned as code Alpha for the control group and Bravo for custody children. After a session of interated clustering, the engine worked out six groups, including the original Alpha and Bravo. The new groups were arbitrarily named Charlie, Delta, Echo and Foxtrot. This test was a preliminary probe to determine what kind of information can be inferred, and how it can be reported to consist a decision support. The data was automatically assembled into radar graphs by the engine using matplotlib (Hunter, 2007). The primary results are shown in the charts below:

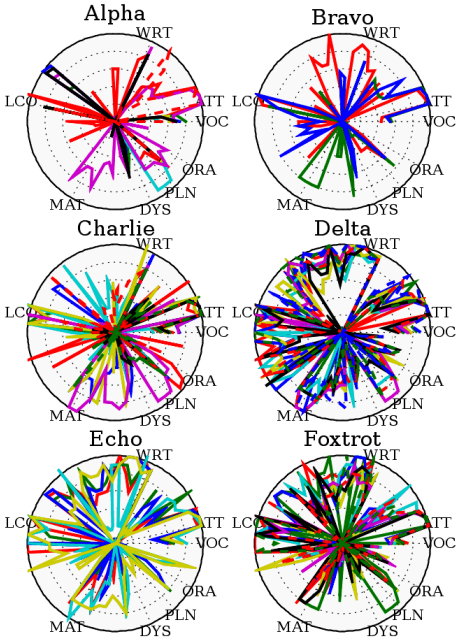


Figure 3: Dysfunction distribution graph.

The charts show the strength and weakness areas as crests and dents. This charts can help to decide which children needs the most urgent handling and which areas are more affected. The labels represent Vocalization, Attention, Writing, Language, Mathematics, Dyslexia, Planning and Orality. Critically damaged Alpha and Bravo children need urgent care and should be highly prioritized in a special learning program, with most areas affected. Echo and Foxtrot children are less impaired but show different education needs with different gaps between weak and strong areas. Language is in need in both and Foxtrot needs more in mathematics and orality areas while Echo have traces of dyslexia. Delta is strong in most areas, while Charlie being good in all logical areas, is rather poor in literacy in general. As a testbed, this experiment gives a good support to the idea of integrating medicals protocols in a computing system. The children classification can help devising a whole overall strategy in dealing with each group, matching the children with its needs.

These reports will be available on line and can be requested as a query on a collection of children. This chart is a sample report with bayesian clustering, but other reports can be devised. The query leaves a formated file in disk and the filename is passed to a program that analyses it and leaves in return the result chart to be displayed. The system uses a framework similar to another scientific web application, called Enviair (Mota *et al*, 2007) developed by our team for environment control.

6 CONCLUSIONS

Children may undergo unnecessary suffering when indications of their actual conditions are neglected, even after all health care protocols have been already applied to them. This is the most ubiquitous cases, where the concerning authorities are not prompted with the proper information, or even if the information is provided, it is hard to roam across a large amount of raw data to figure out what are the proper measures in each case.

This work proposes a qualitative improvement in the programs of prevention, teaching and, in the therapeutic actions and assistances. A web system provides a tool to validate and implement methodologies of technical and administrative intervention. The TUIA system can maintain an extensive collection of integrated data on endangered children, from a wide range of concerning areas. The system not only provides for immediate communication of children condition alerts among the the concerned parts, but also can aggregate relevant and validated research results to the benefit of community. This can be achieved as the system is a platform to develop and validate new protocols and analytical procedures. The experimental protocols can be applied together with the mainstream procedures and the results compared with cross checking from existing data. New protocols and reports can be then incorporated into mainstream as the results have scientific relevance.

The system is developed using state of art web technology to provide fast and consistent development. A Domain-Driven (Evans, 2004) approach is applied to shorten the turnaround of new software releases as new requirements come in to play to extend the system functionality. A lightweight interprocess communication technology supports distributed hosting if required. It integrates with machine learning engines and other advanced analytical tools to provide support to machine assisted inferences on collected data.

The challenge is to congregate the two aspects of the public power role: to create methodologies of scientific investigation, inside the services and in the universities, capturing the complexity of the theme and to incorporate the actions against violence in the involved professionals' practices and to articulate those practices, not only in the section of Health, but also in the Education practices, Social Attendance and Justice.

In a general way, this work proposes the development of an "algorithm" composed by taxonomy of information transformed in analysis methods. The observed population is passed through clustering algorithms, leading to the identification of children's sub-groups, according to patterns in input profiles. This discovery opens the possibility of multiple studies accomplishment that will result in a rich knowledge on abuse and its co-morbidities that do intersection with the several abuse types and classes.

REFERENCES

Beers SR, De Bellis MD. *Neuropsychological function in children with maltreatment-related posttraumatic stress disorder.* Am J Psychiatry. 2002 Mar;159(3):483-6.

Binder, D.A. *Bayesian cluster analysis.* Biometrika 1978 65(1):31-38

De Bellis MD. *The psychobiology of neglect.*
Child Maltreat. 2005 May;10(2):150-72. Review.

DOM - *Document Object Model.* Avaiable at http://www.w3.org/DOM/ (accessed July 15, 2007).

Evans, E. *Domain-Driven Design: Tackling Complexity in the Heart of Software.* Addison-Wesley, Boston, 2004.

Good CD, Johnsrude IS, Ashburner J, Henson RN, Friston KJ, Frackowiak RS. *A voxel-based morphometric study of ageing in 465 normal adult human brains.* Neuroimage. 2001 Jul;14(1 Pt 1):21-36.

Hunter, J. D. *Matplotlib: A 2D Graphics Environment.Computing in Science and Engineering,* vol. 9, no. 3, pp. 90-95, May/Jun, 2007.

JSON - *JavaScript Object Notation*. Avaiable at http://www.json.org/ (accessed June 9, 2007).

Khare, R., Taylor, R.N. *Extending the Representational State Transfer (REST) architectural style for decentralized systems.* ICSE 2004. Proceedings. 26th International Conference on Software Engineering, 2004.

Meyer-Lindenberg A, Zink CF. *Imaging Genetics for Neuropsychiatric Disorders.* Child Adolesc Psychiatr Clin N Am. 2007 Jul;16(3):581-597.

Moll J, Zahn R, de Oliveira-Souza R, Krueger F, Grafman J. *Opinion: the neural basis of human moral cognition.* Nat Rev Neurosci. 2005 Oct;6(10):799-809.

Mota, L.C; Oliveira, C.E.T; Meirelles, M.S.P; Berroir, J.P.; Herlin, I. *The Implementation of a web-based System for Automatic Classification of land use and covering changes.* IADIS International Conference Applied Computing: Salamanca, 2007.

Tovar-Moll F, Moll J, de Oliveira-Souza R, Bramati I, Andreiuolo PA, Lent R. *Neuroplasticity in human callosal dysgenesis: a diffusion tensor imaging study.* Cereb Cortex. 2007 Mar;17(3):531-41.

Sternberg KJ, Knutson JF, Lamb ME, Baradaran LP, Nolan CM, Flanzer S. *The child maltreatment log: a computer-based program for describing research samples.* Child Maltreat. 2004 Feb;9(1):30-48.

AN ONTOLOGY-BASED INFORMATION SYSTEM FOR MULTICENTER EPIDEMIOLOGIC STUDIES ON CANCER

J. M. Vázquez[1], M. Martínez[1], M. G. López[1], B. González-Conde[2], F. M. Arnal[2]
J. Pereira[1] and A. Pazos[1]
[1]*Medical Informatics and Radiological Diagnosis Center (IMEDIR), University of A Coruña, A Coruña, Spain*
[2]*Universitary Hospital Complex Juan Canalejo (CHUJC), A Coruña, Spain*
jmvazquez@udc.es, marcosmartinez@udc.es, mglopez@udc.es,
bgoncon@canalejo.org, farnmon@canalejo.org, javierp@udc.es, apazos@udc.es

Keywords: Cancer, information systems, ontologies, epidemiology, handheld devices, security.

Abstract: Diseases like cancer are caused by a diversity of different factors interacting together, whose study requires a huge amount of data. Compiling this data is an expensive and time-consuming task that can be carried out in an easier, faster and more secure way with the support of Information and Communication Technologies (ICT). Nevertheless, the majority of epidemiologic studies are executed without this support of informatics or with basic tools that are developed by unqualified professionals. As a consequence, the integrity of the collected data cannot be assured, and the reliability of the studies is usually decreased. This work presents an ontology-based Information System for the development of multicenter epidemiologic studies on cancer that allows 1) collecting, storing and editing medical data from different hospitals and 2) reusing the compiled data by means of their integration with data from other systems. This system has been satisfactorily applied to an epidemiologic study of colorectal cancer in Galicia, Spain.

1 INTRODUCTION

Epidemiologic studies are useful to understand the origin of diseases, to detect outbreaks of pathologies in the population, to make decisions to optimize resources and, in sum, to improve the welfare and quality of life for societies worldwide. However, the study of multifactorial diseases like cancer, which are caused by a variety of genetic, environmental and lifestyle factors requires a large amount of data.

Compiling these data is a laborious work that implies: a) the recovery of the patient's clinical records, b) the recovery of analysis data, and c) personal interviews with the patient and his family in order to fill out various questionnaires (e.g. family questionnaires, risk factor questionnaires, etc.). In addition, medical personnel in hospitals is usually under a lot of attendance pressure and it is very difficult for them to devote time and energy to arduous tasks of interviews and information collection. Therefore, facilitating as far as possible this work becomes a critical task.

Information and Communication Technologies (ICT) allow the development of Information Systems (ISs) that provide mechanisms for the data collection (including validation rules and error control), as well as the storage and editing of medical data from various hospitals by means of the Internet. ICT allow us to efficiently manage large amounts of data and therefore enhance the quality of epidemiologic studies. It is also possible to use ICT and the Internet to integrate various data sources and as such obtain an even larger set of data.

In spite of these obvious advances, there still exists a problem that concerns a great part of epidemiologic studies: they are usually either carried out manually, or using basic tools that have been developed by unqualified personnel (it is only during the data analysis phase when professional tools tend to be used). Moreover, these tools are rarely designed to allow their integration with other systems. In consequence, the integrity of the collected data cannot be guaranteed and the reliability of the studies is frequently decreased.

2 BACKGROUND

The region of Galicia, situated in the northwest of Spain, represents an excellent geographical area for carrying out genetic-epidemiologic studies of colon and rectal cancer due to the homogeneity of its population in several dimensions: genetically (which facilitates this type of studies) as well as culturally and environmentally (which allows for conducting homogeneous recruiting subjects that would participate in the study, as well as providing an opportunity to study gene-environment interactions). In addition, there are other, equally important factors associated with conducting such a study in Galicia, such as the relatively high incidence of colorectal cancer in its population, and the availability of subject families (which are frequently large families that live in the same city or town).

All of these factors have induced the development of several recent research projects on cancer in Galicia during the last years: "A Pilot Study of colorectal cancer in Galicia, Spain", funded by the U. S. National Cancer Institute (NCI) for the period 2004-2006, a "Colorectal Cancer Thematic Network in Galicia", funded by the *XUNTA de Galicia* for the period 2005-2006 and a "Colorectal Cancer Research Network in Galicia", funded by the *XUNTA de Galicia* for the period 2006-2008/9. Within the framework of these projects, the idea of developing an IS to improve the development of multicenter epidemiologic studies, was jointly raised in 2005 by the Universitary Hospital Complex Juan Canalejo of A Coruña and the Medical Computing and Radiological Diagnosis Center (IMEDIR Center) of the University of A Coruña, and funded by the *XUNTA de Galicia* for the period 2005-2008.

3 OBJECTIVES

The aim of the this work consists in developing a secure IS for the achievement of multicenter epidemiologic studies on cancer that allows 1) collecting, storing and editing medical data from different hospital centers, and 2) reusing the compiled data, by means of their integration with data from other systems in order to carry out studies on a larger set of data.

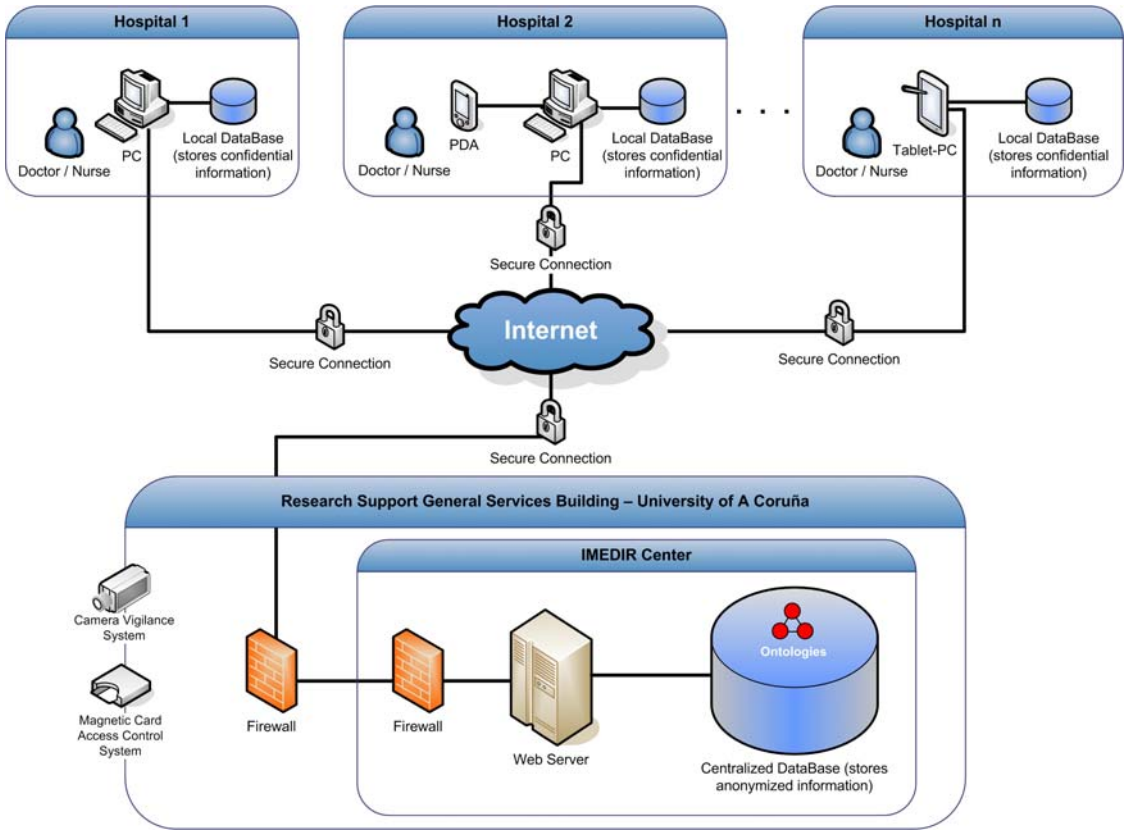


Figure 1: Physical design of the Information System.

4 METHODS

The development of the system was based on the Unified Software Development Process (USDP, Jacobson, 1999). During the Software Requirements Specification phase (SRS), we used prototyping techniques and interviewed many medical experts in different hospitals. This allowed us to detect with great detail the system requirements and to define the interfaces according to the preferences of the end-users.

The central system, located in the IMEDIR Center, was developed on the J2EE platform in order to offer more integration possibilities and follows the architectonic patterns Model-View-Controller (MVC) and Layers. The software for the data compilation devices used in hospitals was developed using the .NET Compact Framework, which simplifies application development on smart devices and allows to develop a very friendly user interface. The physical design of the system is shown in the Figure 1.

4.1 Distributed Data Collection

We studied and discussed various data collection alternatives (Tablet PCs, PDAs, Smart Phones) and finally opted for using Portable Digital Assistants (PDAs), due to their characteristics of mobility, data synchronization (e.g. with a desktop computer), data input facilities (pen-stylus method) and an enough size screen for a correct handling of the application (Wiggins, 2004). The software for the PDAs was developed on the .NET Compact Framework because it makes it easier to build applications for this kind of handheld devices and allows to develop a very friendly user interface (see Figure 2). We must consider that usability is a fundamental factor to obtain the success of an IS, especially in medical environments. Although at present only data compilation software for PDAs is available, the system is independent from the data collection device used, so it can be easily adapted to other devices like the previously mentioned.

The data collection software allows the automatic validation of data, minimizing the input errors (e.g. personal names cannot contain numerical characters, birth dates must be past dates, etc.). In addition, this software is able to avoid asking questions whose answers can be inferred from previous answers (e.g. if a patient have already provided his birth date, he will not have to fill out his age) and it also provides assistance to fill out the required information: it has an on-screen help system, shows examples of possible data inputs in confusing questions, guides the user (e.g. the person who makes the interview) through the questions (because in some questionnaires the user must follow a way or another one depending on previous answers), avoids unasking questions by mistake or omission, etc. All these features have been implemented in order to decrease the length of the medical interviews and, therefore, to reduce the cost of the studies.

The data collected from hospitals using the PDAs are formatted to XML and synchronized at regular intervals by a device (PC) that is located in each hospital and connected with the central system at IMEDIR Center through a secure Internet connection. A Web application, which was developed using J2EE technology, allows the transmission and storage of the collected data from the PCs in hospitals to the centralized database.

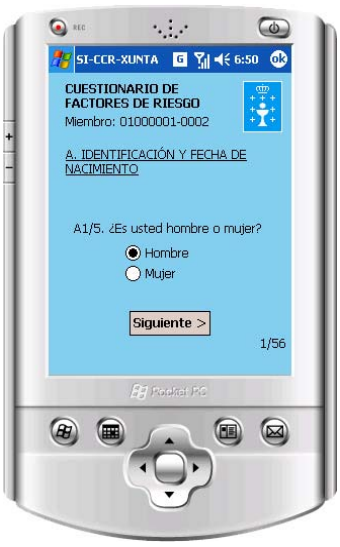


Figure 2: Screenshot of the data collection software.

4.2 Centralized Data Storage

In order to maintain the data consistency, which is fundamental in order that the studies would be valid, and to facilitate the data exploitation, the IS stores all the collected data in a centralized database.

Since the proposed system stores information arriving from multiple hospitals, which could be placed in different geographical areas (different cities or even in different countries), and due to the fact that the collected data belongs to the Medical domain, in which there exists a great terminological

heterogeneity, it is crucial to use a standard terminology for the data storage.

At present, ontologies are viewed as an ideal solution to solve data heterogeneity problems. They are solid vocabularies of terms and relations among them, agreed upon by a group of people, that can help to overcome the semantic, syntactic and structural ambiguity that hinders communication between different systems and data sources.

After analyzing several options, we decided to use the NCI Thesaurus ontology in our system (Golbeck et al., 2004), because it is published under an open content license and it contains a broader range of cancer-related terms than other existing ontologies. Using this ontology as a reference, the information coming from each hospital is annotated by means of a common terminology in which it is stored. This allows the use of questionnaires that can be written in different languages or medical terminologies depending on the geographical location of each hospital, because the storage is done in agreement with a common terminology. Furthermore, this process is not restricted to the translation of terms, but it also covers other aspects like, for example, units of measurement or date and hour formats.

The system is flexible enough to use other ontology instead of the NCI Thesaurus; however, it only supports using one ontology at a time.

4.3 Remote Data Editing

In the proposed system, data editing refers to the act of modifying or deleting incorrect information from patients and/or their relatives, or storing new information that was not known before. This may be required by the medical staff 1) before the data have been transmitted from the hospital to the centralized database, or 2) when the data are already stored in the centralized database. In the following, both kinds of data edition are described.

The first case lies in the edition of data that are still stored in the hospitals. They are data that have not been transmitted to the centralized database because they are incomplete, or because they have not been reviewed yet. The user must authenticate himself to the system and carry out the appropriate changes. This is a usual kind of data edition.

The second case refers to the edition of data that have already been sent from the hospital to the centralized database. In this case, the user works directly against the centralized database. This option should be used only in exceptional cases, because it consists of modifying information that is already assumed to be complete and reviewed. As a safety measure that prevents the system about the loss of important information due to user errors or about possible attacks from outside, this kind of edition requires, in addition to the authentication of the doctor or nurse who wishes to edit the data, the permission of the system administrator, who is located in the IMEDIR Center.

4.4 Data Integration Capabilities

The data collected during the achievement of an epidemiologic study has a great value, both because of the difficulty and cost (in time and economic) of this process and due to its great potential of reusability by means of its integration with data from other sources, which allows to carry out new studies with a larger amount of data.

Nevertheless, the data gathered during an epidemiologic study are rarely reused after its finalization. This is mainly due to that traditional storage supports (e.g. paper) are used, as well as specific storage formats and terminologies which cause that the reusability of the collected data by means of the integration with data coming from other studies is not worthwhile.

By means of the last advances in ICT, we have provided our system with capabilities of information reusing and integration.

The main objective of data integration is to provide ways to unify the information from several distributed, heterogeneous and autonomous data sources (e.g. information systems, databases, XML files, etc.). An integrated view must be able to describe the various data sources and their interrelation, overcoming the syntactic, structural and semantic heterogeneity problems. All of this, with the aim of automating the process of getting data from various resources, instead of having to manually request data from them and then combine the results.

However, as it is explained in Chou, 2005, Information integration is not a trivial task. Data are usually stored in relational databases, and it is often the case that only the creators of the databases understand the semantic meaning of columns in each table. Therefore, it is difficult for a user (or a system) to integrate the data from the databases with data from other sources without first understanding how they are structured, or without being explicitly told from which columns to retrieve information.

The presented system allows 1) accessing through the Internet to the information collected during an epidemiologic study on colorectal cancer

in Galicia, Spain and 2) requesting information from any set of information sources which have been mapped to an ontology from the cancer domain. In the following, we will describe how these two functionalities were achieved.

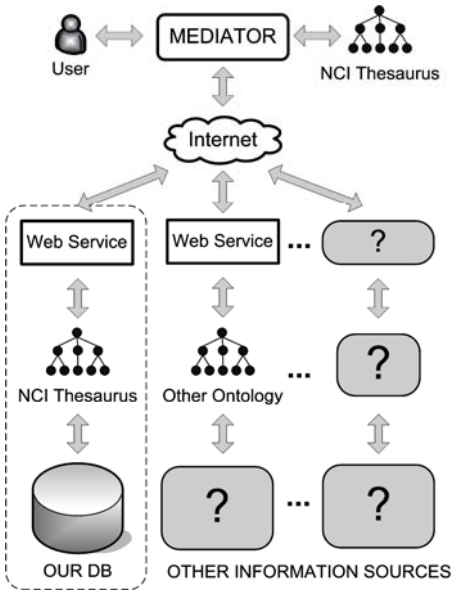


Figure 3: Data integration capabilities of the IS.

4.4.1 Making the Collected Data Publicly Accessible

In order to make our information publicly accessible through the Internet, we have opted for a solution that uses ontologies and Web services.

Ontologies are useful to support the integration of data from multiple repositories (Jakoniene & Lambrix, 2005, Perez-Rey et al., 2005, Stevens et al., 2000), and some of the current Integration Information Systems incorporate ontology-related knowledge (e.g. Deray & Verheyden, 2003, Alexiev et al. 2005). The proposed system uses ontologies to resolve the semantic conflicts that usually hinder integration data by using the NCI Thesaurus ontology as a reference vocabulary, mapping the columns of the centralized database with terms from that ontology.

On the other hand, Web services provide loosely-coupled, language-neutral, and platform-independent ways of linking applications across the Internet. Our system is designed to allow remote queries written in the terminology of the NCI Thesaurus ontology, and to answer these queries through the Internet. The elements of the IS that provide this functionality are represented in the Figure 3 into a dotted area.

4.4.2 Integrating Information from other Sources

Instead of having to manually request data from various data sources and then combine the results, our system is also prepared to automate this process. We have implemented this functionality on the basis of one of the ontology-based information integration approaches (the "Single Ontology Approach"), proposed by Wache et al. in 2001.

In this process (see Figure 3), the IS acts as a "Mediator" that 1) receives a request (query) from the user, 2) processes the query and ask a set of data sources that have been prepared to be accessed through Web services, and 3) puts together all the results from the data sources and returns the combined result to the user.

At the moment, this functionality is in testing phase, and it only works well with information sources that are available to be queried through a Web service by means of the terminology of the NCI Thesaurus. However, the preliminary results have been very satisfactory, and our intention is to continue improving this functionality to reach the most general and automatic behaviour possible.

4.5 Security Issues

The development of this IS also considers the security requirements imposed by Spanish law and Galician regulations, which are among the most restrictive European legislations, as well as the United States Safe Harbor Agreement. Under the Spanish legislation in force, the medical data in this IS are considered sensible data and specially protected, thus the safety measures acquire special importance. The data collection is made by the medical staff in hospitals, that is authorized by the law for the managing and processing of data about personal health (LOPD 15/1999, 1999). The used PDAs have an integrated biometric fingerprint reader, which provides security access to the personal data stored in the device and univocally identifies the user, according to the measures arranged by the law (RD 994/1999 , 1999).

In order to transfer the collected data from the PDAs to the PC, both devices are connected by cable, and the doctor or nurse is authenticated in the PC by means of a cryptographic smart card of the Galician Service of Health (SERGAS). In this smart card there is stored a digital certificate issued by the *Fábrica Nacional de Moneda y Timbre* (FNMT), a certifier authority recognized by the Spanish state that univocally identifies the user who uses it. All

accesses to the data are totally monitored and registered, and it is registered for the later accomplishment of audits (Wei et al., 2006). When transferring the data to the PC, a dissociation process is made in which the personal character data necessary for the medical personnel to identify the subject (e.g. the number of clinical history of the patient) and the genetic-environmental data required by the epidemiologic study, that will be transferred later to the centralized database located in the IMEDIR Center, are separated. In this process the patient's identity is dissociated of its clinical data, which are anonimous under a numerical code, and both are stored in the PC in a separated way. The relation between both data types is stored in a file which will only be accessed from the hospital.

The transmission of the anonimous data from the PCs in the hospitals to the centralized database is made over the Internet and through a Web application, in which the user is authenticated by means of the same mechanism that he/she uses when is connected to the PC in the hospital, that is, by means of the digital certificate that resides in the cryptographic smart card of the SERGAS. To assure the safety of the data during the transmission, all the data tranfers are carried out on encrypted connections using the Secure HyperText Transfer Protocol (HTTPS) over the Secure Socket Layer (SSL). To incorporate this method, a server security certificate needs to be configured on the server, so these technologies and protocols use public/private key technologies (Cooper et al., 2006, Bourasa et al., 2005). Likewise, the IMEDIR Center has an architecture of double firewall (see Figure 1), in which the first firewall of the building limits the access to prevent external generic attacks, whereas the second firewall, placed inside the IMEDIR Center, restricts the access to the Web application by IP address, so that only those IPs that have been authorized (the PCs of the hospitals) can connect with the application to transfer the data. This system has been chosen as the most suitable due to the fact that resting on the HTTPS protocol the development of Web services is quite simple, and they can take advantage of the firewall safety systems without need to change the filter rules (Stanton, 2005).

Although the data stored in the IMEDIR Center are anonymous, the peculiar characteristics of some gathered families might allow their identification, what makes necessary to maintain a high level of security at all time. With the purpose of providing an environment as safe as possible in the IMEDIR Center, it has several physical safety measures, like security cameras that provide 24-hour video vigilance and the use of cryptographic smart cards to control the access to the building.

5 RESULTS

The presented IS allows to carry out epidemiologic studies on cancer that require less time for interviews with patients and present less errors in the compiled data, while guaranteeing the integrity of these data. The system satisfies the special demands of modern medical information systems, such as security and interoperability. The use of this IS allows to save time and money, and increase the reliability of the performed studies. It also allows us to integrate our data with other information systems, through the Internet, and as such carry out new studies with a larger amount of data (this is particularly important in cancer studies).

This system has been successfully applied in the execution of the "Pilot Study of Colorectal Cancer in Galicia, Spain", financed by the U.S. National Cancer Institute.

6 CONCLUSIONS

The study of multifactorial diseases such as cancer requires a large amount of data that need to be compiled, stored and analysed, and from which new information must be extracted. In addition, reusing these data in other similar studies would provide great benefits.

Information and Communication Technologies can contribute significantly to this task thanks to the development of Information Systems such as the presently proposed one. This system, allows collecting, storing and editing medical data from different hospitals in a secure manner, and reusing the compiled data by means of their integration with data from other information sources with the purpose of carrying out studies on a larger set of data. The usefulness of this IS has been demonstrated during the development of a real epidemiologic study of colorectal cancer in Galicia, Spain.

7 FUTURE DIRECTIONS

In the following, some of the ideas that could help to improve the proposed IS are presented:

With regard to the data integration capabilities of the system, we are thinking about developing an advanced mechanism to automatically retrieving information from sources whose information has been prepared to be accessed. This mechanism could be based on a set of intelligent semantic agents that would interoperate with the various information sources through the terminology of existing ontologies. This would allow us automatically retrieve and integrate a huge amount of data from other studies that we would analyse in order to make progresses in the treatment of pathologies like colorectal cancer.

It also could be useful to provide the system with data mining techniques (Tan, Steinbach & Kumar, 2006). These techniques could be deployed to scour the large amount of epidemiologic data compiled in order to find novel and useful patterns that might otherwise remain unknown, and they would also be useful to predict the outcome of future observations. All this would help to decrease the incidence of diseases like the cancer, and to improve its prevention and treatment.

ACKNOWLEDGEMENTS

This work was partially supported by the Spanish Ministry of Education and Culture (Ref TIN2006-13274) and the European Regional Development Funds (ERDF), grant (Ref. PIO52048) funded by the Carlos III Health Institute, grant (Ref. PGIDIT 05 SIN 10501PR) from the General Directorate of Research of the Xunta de Galicia and grant (File 2006/60) from the General Directorate of Scientific and Technologic Promotion of the Galician University System of the Xunta de Galicia. The work of José M. Vázquez is supported by an FPU grant (Ref. AP2005-1415) from the Spanish Ministry of Education and Science.

REFERENCES

Alexiev, V., Breu, M., de Bruijn, J., Fensel, D., Lara, R., Lausen H., 2005. *Information Integration with Ontologies – Experiences from an Industrial Showcase*. John Wiley & Sons.

Bourasa, C., Gkamasa, A., Naveb, I., Primpasa, D., Shanib, A., Sheoryb, O., Stamosa, K., Tzruyac, Y., 2005. Application on demand system over the Internet. *Journal of Network and Computer Applications*. 28(3), 209-232.

Cooper, C. J., Cooper, S. P., Junco, D. J., Shipp, E. M., Whitworth, R., Cooper S. R., 2006. Web-based data collection: detailed methods of a questionnaire and data gathering tool. *Epidemiologic Perspectives & Innovations*. Volume 3.

Chou, H. H., 2005. *BioDig: Architecture for Integrating Heterogeneous Biological Data Repositories Using Ontologies*. Thesis. Massachusetts Institute of Technology.

Deray, T., Verheyden, P., 2003. *Towards a Semantic Integration of Medical Relational Databases by Using Ontologies: A Case Study*. OTM Workshops, LNCS 2889, 137-150. Springer-Verlag.

Golbeck, J., Fragoso, G., Hartel, F., Hendel, J., Parsia, B., 2004. The National Cancer Institute's Thesaurus and Ontology. *Journal of Web Semantics*, Vol. 1, issue 1.

Jacobson, I., Booch, G., Rumbaugh, J., 1999. *The Unified Software Development Process*. Addison Wesley.

Jakoniene, V., and Lambrix, P., 2005. Ontology-based Integration for Bioinformatics. *Proceedings of VLDB Workshop on Ontologies-based techniques for DataBases and Information Systems (ODBIS2005)*, Trondheim, Norway.

Organic Law 15/1999 of December 13, 1999 on the Protection of Personal Data.

Royal Decree 994/1999 of June 11, 1999.

Perez-Rey, D., Maojo, V., Garcia-Remesal, M., Alonso-Calvo, R., Billhardt, H., Martin-Sanchez, F., Sousa, A., 2005. ONTOFUSION: Ontology-Based Integration of Genomic and Clinical Databases. *Computers in Biology and Medicine*.

Stanton, R., 2005. Securing VPNs: comparing SSL and IPsec. *Computer Fraud & Security*. 2005(9), 17-19.

Stevens, R., Baker, P., Bechhofer, S., Ng, G., Jacoby, A., Paton, N., Goble, C., Brass, A., 2000. TAMBIS: Transparent access to multiple bioinformatics information sources. *Bioinformatics*, 16(2), 184-186.

Tan, P., Steinbach, M. and Kumar, V., 2006. Introduction to Data Mining. Addison Wesley.

Wache, H., Vogele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., Hiibner, S., 2001. Ontology-based integration of information - a survey of existing approaches. *Proceedings of the International Workshop on Ontologies and Information Sharing*, 108-117.

Wei, J. C., Valentino, D. J., Bell, D. S., Baker, R. S., 2006. A Web-based telemedicine system for diabetic retinopathy screening using digital fundus photography. *Telemed J E Health*, 12(1):50-7.

Wiggins, R.H., 2004. Personal digital assistants. *Journal of Digital Imaging*. 17(1):5-17.

ADAPTIVE CLINICAL PATHWAYS WITH SEMANTIC WEB RULES

Dimitrios Alexandrou, Fotis Xenikoudakis and Gregoris Mentzas
Information Management Unit, Electrical and Computer Engineering, National Technical University of Athens
9 Iroon Politechniou Street, Athens, Greece
dalexan@mail.ntua.gr , fxeni@image.ece.ntua.gr, gmentzas@mail.ntua.gr

Keywords: Adaptive clinical pathways, semantic web, semantic rules, pathway adaptation, ontology, workflow engine.

Abstract: The increase of treatment quality offered by the healthcare organizations is one of the main challenges of the modern health informatics. The personalization of treatment presupposes the real-time adaptation of treatment schemes since the clinical status of the patient and circumstances inside a healthcare organization constantly change. In this paper we present SEMPATH prototype which aims at providing a solution concerning the real-time adaptation of healthcare business processes. The prototype consists of a healthcare process execution engine assisted by a semantic framework for the adaptation. The semantic framework consists of an ontology enclosing the required knowledge based on which a semantic rule set was created. During the execution time of the clinical pathways, the system reasons over the rules, the knowledge and information collected, and provides decisions and recommendations for the next steps of the treatment. Moreover, the results of the rule-set execution may produce new knowledge objects which are inserted in the ontology.

1 INTRODUCTION

One of the main challenges of modern healthcare organizations is to increase the treatment quality. In order to achieve their goal, they need to utilize standardized clinical protocols used in many domains of medicine. Such a protocol contains detailed medical plans for diagnosis, therapy scheme and follow-up. Moreover, it encloses the information required so as to deal with exceptional situations, which occur during the treatment execution time and require quick and appropriate modifications of the treatment of a patient, thus increasing the flexibility of the treatment processes. One valuable tool to achieve the above-mentioned objectives is "Clinical Pathways".

Clinical pathways can be utilized for the implementation of medical guidelines in a specific healthcare environment and decrease undesired variability of medical practice (Campbell et al., 1998). In contradiction with the medical guidelines, clinical pathways enclose multidisciplinary valuable resources like personnel, education level, medical equipment availability and other operational and administrative information. Medical guidelines require the consensus between medical experts. On the other hand, clinical pathways require a consensus between multidisciplinary groups of hospital personnel taking actions during the treatment execution. Clinical pathways constitute treatment process patterns which aim to increase both the healthcare process quality and the utilization of resources. Consequently, a clinical pathway may deviate from a clinical guideline due to administrative reasons, and a treatment scheme may deviate from the clinical pathway due to patient's symptoms during its execution.

In order to support the execution of treatment schemes based on clinical pathywas and to relief the medical personnel, a software system is required which will handle the healthcare business processes in an efficient manner (Greiner et al., 2004). Such a system would be responsible for the observation of the execution and the current status of the apllied clinical pathways, offer the characteristic of automatic recognition of exceptional events and provide decision support services in order to handle the exceptions in an efficient and effective way. Moreover, the software system should be capable to dynamically adapt the treatment process so as to control the appropriate modifications.

In this paper we propose an approach which includes a workflow management system combined with a rule base in order to handle the abovementioned requirements. The workflow environment handles the execution of treatment schemes and the incorporation of user types, data and peripheral applications. Additionally, the specific software system needs to support the dynamic adaptation of clinical pathways in order to handle the flexibility of treatment schemes (Dadam et al., 2000), (Miksch et al., 2001). The rule base is responsible for the handling of the required streams of knowledge enclosed in the clinical pathways and is utilized for the detection of exceptional events and their confrontation.

In this paper we present our software prototype, SEMPATH, which follows this approach and has all the required functionality to support adaptive clinical pathways. SEMPATH performs a rule-based exception detection with semantic rules (SWRL) and dynamic clinical pathway adaptation during the execution time of each pathway.

The rest of the paper is organised as follows. Section 2 refers to our motivations and related work performed in the area of our interest. Section 3 overviews the proposed Semantic Approach the SEMPATH follows, while Section 4 outlines the SEMPATH conceptual framework and technical architecture which is being implemented. In section 5 we present our experimental scenario and the SEMPATH walkthrough. Finally, section 6 concludes the paper combined with our thoughts for future work.

2 MOTIVATIONS AND RELATED WORK

At this section we present the motivations, the related work and our contribution in the area of the adaptive clinical pathways. The motivations presented led to the research stream of adaptive clinical pathways. Moreover, a significant amount of work has been realized towards the direction of the optimal handling of the exceptions occurring during the execution of treatment schemes of a patient. Finally, our research in the area and the development of SEMPATH prototype tries to contribute in specific and focused issues.

2.1 Motivations

The trends in healthcare business processes and their establishment and utilization in the healthcare routine are up to now quite mature. Nevertheless, there are several open issues / challenges that further motivate our effort (Song et al., 2006):

- **Clinical Pathways Adaptability:** The traditional clinical pathways are normally static and lack of dynamicity. Moreover, they are standard procedures applicable to a patient taxonomy not addressing the case of each patient. Moreover, they do not take under consideration the most current medical, operational, and financial knowledge (Colaert, 2007).
- **Maintenance:** The implementation of Clinical Pathways is based on medical guidelines and additional types of knowledge. The maintenance of the healthcare business process suffers from the continuous update, since both the medical guidelines and the circumstances inside a healthcare organization change constantly.
- **Medical Guidelines Formalization:** The formalization of medical guidelines is being performed in a specific and per case manner. Their formalization is required since their parameters will be able to be processed by an IT infrastructure that supports their execution.
- **Clinical Pathways Modelling:** The modelling of Clinical Pathways lacks a formal structure. Different approaches exist in the area of modelling. Their interoperation could be of major importance since the Clinical Pathway exchange between healthcare organizations could facilitate the execution of the treatment schemes utilized.
- **Real-time information capturing:** Information capturing consists one of the major factors for success of the treatment scheme executed for each patient. The lack of real-time information feed to the clinical pathway creates a major need, since the information collected could lead to major reconfigurations of the executed Clinical Pathway.
- **Real-time knowledge recycling:** The knowledge recycling during the execution of a Clinical Pathway constitutes one of the major challenges for the area. The knowledge feedback would be valuable since the knowledge update is able to redefine the Clinical Pathway and the model of the exception rules.

2.2 Related Work – State of the Art

As (Lenz et al., 2006) states, "*healthcare processes require interdisciplinary cooperation and coordination*". Towards this direction, he divides the processes inside a healthcare organization into two categories: the organizational processes and the medical treatment processes. The organizational processes are of equal importance to the medical treatment ones, since they heavily affect their execution and effectiveness. Moreover, the medical treatment processes are influenced by the medical knowledge and the patient information. So, he introduces the need for WfMS (Workflow Management System) inside a healthcare organization so as to handle the intra-organizational processes. Moreover, the addition of the appropriate web-services could lead to the inter-organizational healthcare processes. The abovementioned concerns led to the implementation of ADEPT system (Reichert et al., 2003) which focuses on the healthcare processes execution. The ADEPT system enables the execution, monitoring and management of the healthcare process running inside a healthcare organization. Moreover, it offers the functionality of dynamic changes in the predefined healthcare processes on execution time. The development of the specific system lasted for some years and provided valuable information and experience from its pilot and productive periods (Lenz et al., 2007), (Blaser et al., 2007).

Additionally, (Colaert, D., 2007) introduces the term of adaptive clinical pathways and presents the research work performed inside Agfa Healthcare. He stresses out that Adaptive Clinical Workflows are based on a) Medical, b) Practice, c) Clinical and d) Operational Knowledge. Agfa constitutes one of the active members of W3C Semantic Web Health Care and Life Sciences Interest Group (W3C) which encloses the "Adaptive Healthcare Protocols and Pathways Task Force" which aims at the utilization of semantic web technologies in order to enhance the adaptable clinical protocols and pathways.

(Abidi and Chen, 2006) introduce another IT platform that enables the adaptivity of clinical pathways based on a semantic framework. CAREPLAN system (Abidi and Chen, 2006) tries to combine heterogeneous healthcare knowledge sources with the available patient information. The system reasons over the knowledge and adapts standard pathways towards personalized healthcare plans, utilizing the technology of web-services for the composition of the integrated pathways.

2.3 Our Contribution

Our approach led to the creation of the SEMPATH software system which enables the adaptation of clinical pathways in order to serve the personalization of the treatment plans for each patient. Our contribution concerning the state-of-the-art in the specific domain could be summarized in the following axes:

- **Real-time adaptation of clinical pathways:** SEMPATH approach is based on continuous reasoning over the current knowledge so as to adapt each step of the clinical pathway under execution.
- **SWRL Rule Base:** SEMPATH encloses a rule-set created by utilizing SWRL (SWRL) language in order to integrate the rule-base with the ontology. The rule-base is able to create new facts and update the ontology accordingly, thus creating new knowledge as each pathway evolves. This feedback constantly updates the knowledge stored in the ontology and leads to better results concerning the adaptation of the pathway.
- **Establishment of a meta-model for each clinical pathway:** in our approach, we define a meta-model for each clinical pathway to be executed. The meta-model encloses atomic and complex sub-pathways which are fed to the process execution engine. The integration of discrete parts and connections results to the establishment of the meta-model of the pathway to be executed.

3 THE SEMPATH APPROACH

The SEMPATH approach followed during the implementation of the software prototype includes the design and creation of the Adaptive Clinical Pathway Ontology. The created ontology is utilized for the implementation of the semantic web rules. The implemented rules are inserted into the JESS System to comprise the rule-base in order to be executed and extract the appropriate facts that influence the adaptation of the clinical pathways.

3.1 Adaptive Clinical Pathway Ontology

The Adaptive Clinical Pathway Ontology constitutes the main infostructure of the semantic layer of the implemented architecture. The core of the ontology is based on the ACPP Ontology (ACPP Ontology)

that is implemented by ACPP Task Force (ACPP Task Force).

The core of the ontology was further extended and broadened both in subdomains and existing concepts and instances. The specific ontology is divided into four (4) main knowledge streams.

The main stream refers to the representation of the medical knowledge. It contains the semantics utilized for the description and structure of the medical part of the clinical pathway. The specific concepts describe the medical domain knowledge that semantically enhances each clinical pathway.

Moreover, the second stream of the ontology comprises the operational knowledge structure. It contains the concepts utilized for the description of the operational issues that arise during the execution of the clinical pathway and may affect its evolvement. Each healthcare organization encloses specific procedures and resources which are combined in order to offer its services. Consequently, the operational knowledge is one of the main elements of a clinical pathway since it affects its execution and success.

Additionally, the third knowledge stream refers to the concepts and terms that define the financial issues that affect the execution of a clinical pathway. The utilization of a clinical pathway aims at the optimization of the financial resources required for the treatment path of a patient. Each healthcare organization aims at both the reduction of its costs and the increment of the quality of healthcare services provided. Consequently, the financial part of the ontology models the financial resources and rules utilized during the execution of each clinical pathway.

Finally, the fourth knowledge stream refers to the modelling of the clinical pathway itself. It contains the concepts and terms that describe the building blocks of the clinical pathway. It is utilized by the software prototype for the design of the treatment workflow.

SEMPATH Ontology is available online at: http://www.imu.iccs.gr/index.php?option=com_content&task=view&id=206&Itemid=90/sempath_onto.zip.

3.2 SWRL Rules Modelling

The rules implemented for our prototype refer to the execution of a clinical pathway, and more specifically to the exception handling procedure. The exception management procedure constitutes one of the every-day routine of healthcare professionals (Kobayashi et al, 2005), (Tucker et al, 2002). As (Han et al., 2006) states, "a*n exception, constitutes an abnormal behaviour from the normal workflow*". The handling of exceptions occurring during the execution of a clinical pathway encloses three major issues: a) exception management representation, b) implementation and execution of exception management and c) exception analysis.

Concerning SEMPATH prototype, the rules for exception management are designed in SWRL format. SWRL enables the integration of the modelled rules with the Clinical Pathway Ontology. The interaction between rules and ontology leads to new knowledge. An indicative set of implemented rules for the SEMPATH prototype is presented below:

Rule-1:

TriageAdmission(?t) ∧
DiagnosedNeurologicalDeficit(?s) ∧
hasTask(?a, ?t) ∧
hasPatient(?a, ?p) ∧
hasPatientState(?p, ?s) ∧
EvaluationForThrombolysisEligibility(?t2)
→ hasNextTask(?a, ?t2)

Rule-1 describes the following situation: if the patient is admitted in the healthcare organization and there is a diagnosis of Neurological Deficit, then the patient has to be evaluated for Thrombolysis eligibility.

Rule-2:

EvaluationForThrombolysisEligibility(?t) ∧
ThrombolysisCandidate(?s) ∧
hasTask(?a, ?t) ∧
hasPatient(?a, ?p) ∧
hasPatientState(?p, ?s) ∧
ConfirmationOfAcuteStroke(?t2)
→ hasNextTask(?a, ?t2)

Rule-2 describes the following situation: if the Evaluation for Thrombolysis Eligibility is confirmed and the patient is Thrombolysis Candidate, then the next examination has to confirm or not the Acute Stroke episode.

Rule-3:

EvaluationForThrombolysisEligibility(?t) ∧

AcuteCVA(?s) ∧
hasTask(?a, ?t) ∧
hasPatient(?a, ?p) ∧
hasPatientState(?p, ?s) ∧
EmergentHeadCT(?t2) ∧
CBC-PT-PTT_stat(?t3)
→ hasNextTask(?a, ?t2) ∧
hasNextTask(?a, ?t3)

Rule-3 describes the following situation: if the Evaluation for Thrombolysis Eligibility is confirmed and the patient's state is Acute CVA, then the patient has to be forwarded to the CT department for Emergent CT and for CBC-PT-PTT examination.

The above-mentioned and explained SWRL rules are indicative elements from rule-base which is utilized by the SEMPATH infostructure. As seen above, the rules contain classes from the ontology, both in the antecedent and the consequent parts of the rules. Any new facts deriving from the rules execution are being added in the ontology as new knowledge objects for future utilization.

3.3 Adaptation Methodology

SEMPATH adaptation methodology is based on a meta-model clinical pathway establishment. Each clinical pathway to be executed is a meta-model of a set of atomic and complex sub-processes. The atomic processes are executable parts of the healthcare business process forwarded to the execution engine. The complex processes are sub-workflows which contain atomic processes and a set of decisions. The atomic and the complex processes are interconnected in the meta-model level. Their connections are based on SWRL rules. Once an atomic or complex process is executed, the rule-base is triggered. The knowledge existing inside the ontology, the current clinical status of the patient and the rule-set are interoperating in order to select the next executable part of the clinical pathway. Thus, the adaptation occurs during each step of the pathway execution.

During each cycle of execution, the triggering of the rule-base may result to new knowledge creation that will be utilized in next steps during the execution. This fact ensures the constant update of medical, organizational and operational knowledge stored inside the ontology and consequently to the rule-base.

4 SEMPATH PROTOTYPE

The following sections present the conceptual framework and technical architecture of the software system prototype that executes the clinical pathways and performs the required dynamic adaptation.

4.1 Conceptual Framework

As depicted in Figure 1, the conceptual framework of the SEMPATH system comprises three (3) distinct architectural layers. The upper layer of the architecture is called "Semantic Layer" since it encloses the required semantic infrastructure.

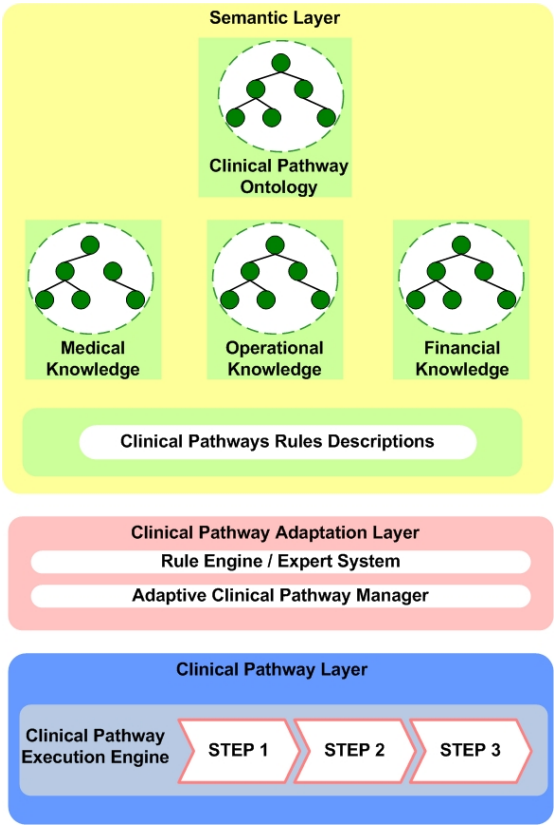


Figure 1: Conceptual Framework.

The core of Semantic Layer is the Adaptive Clinical Pathway Ontology which encloses the appropriate knowledge streams required for the modeling of the clinical pathways, in terms of structure and content. Moreover, the specific layer encloses the semantic modeling of the rules that handle the exceptions inside the clinical pathway during its execution. These rules are the cornerstone of the dynamic adaptation performed in the clinical pathways.

The second layer constitutes the "Clinical Pathway Adaptation Layer". The specific layer encloses the components required for the adaptation of the clinical pathway. The rule engine is responsible for the execution of the semantic rules concerning the exception handling of the pathway. Once the result is produced the "Adaptive Clinical Pathway" manager is the software module that updates accordingly the structure of the pathway.

Finally, the last layer of our conceptual framework is the "Clinical Pathway Layer". It contains the workflow – part of the clinical pathway. The execution of the clinical pathway happens inside a workflow engine since is constitutes a healthcare business process (Lenz, 2006). A clinical pathway repository contains a set of available clinical pathways so as to select the most appropriate for each patient.

4.2 Technical Architecture

The technical architecture of the implemented prototype is presented in Figure 2. The Adaptive Clinical Pathway Prototype technical architecture comprises three (3) major components. These three major components are described in full detail in the following sections:

a. Semantic Infostructure: The core of Semantic Infostructure component is the Clinical Pathway Ontology. As depicted in the following diagram, the ontology is implemented in OWL (OWL) format. The ontology encloses the abovementioned streams of knowledge to be utilized for (1) the creation of rules, (2) the modelling of Clinical Pathways and (3) the recycling of knowledge through the dynamic production of facts by the rule engine.

The Protégé API has been utilized concerning the implementation and maintenance of the specific ontology. Moreover, the SWRL designer of the semantic rules is a Protégé Plugin (SWRLTab) which enhances the integration between the Ontology and SWRL rules. Consequently, each rule created by the specific plugin is consistent in semantic terms, since the semantics required come directly from the ontology.

b. Rule Execution Environment: This component handles the maintenance of the semantic rules as well as the Rule Engine implemented. The SWRL rules implemented with the Protégé Plugin are stored as a SWRL repository. Once the system is triggered, the appropriate rule set is selected. The SWRL rules are initially converted into JESS rules in order to be executed by the rule engine. Once the semantic rules are executed, the result of the rule engine is produced in XML format.

The specific XML file is a custom structure which is utilized by the prototype so as to proceed to the adaptation of the clinical pathway.

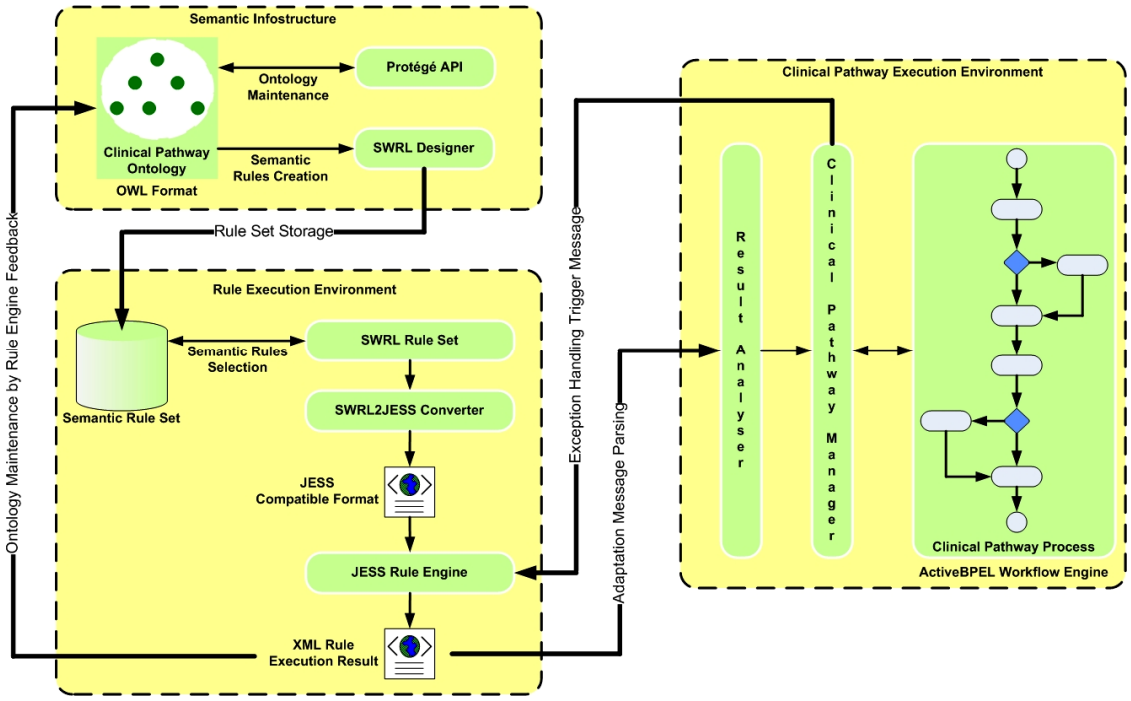


Figure 2: Technical Architecture.

Moreover, once the rule set is executed, despite the production of the XML result, a feedback message is generated which contains new facts and conclusions that update accordingly the knowledge stored inside the ontology.

c. Clinical Pathway Execution Environment: Finally, the last technical component of the prototype architecture is the Execution Environment. The core of the specific component is the workflow execution engine which in our case is ActiveBPEL workflow environment. The interface between the workflow engine and the rest of the components is the Clinical Pathway Manager. Firstly, the specific component triggers the system once an exception occurs during the execution of the pathway. The message produced is forwarded to the Rule Engine in order to run the complete rule-set for the pathway and produce the result for the adaptation. Moreover, the Clinical Pathway Manager interoperates with the Result Analyser which is responsible for the processing of the result structure.

The SEMPATH Prototype Software that we have developed can be found here: http://www.imu.iccs.gr/index.php?option=com_content&task=view&id=206&Itemid=90/sempath.zip.

4.3 Experimental Scenario

According to the real-life scenario, a patient confronts a health problem and decides to be admitted to a healthcare organization for treatment. Once the admission is performed, an initial set of medical examinations is decided to be performed. The result set of the initial examinations provides valuable information for the clinical status of the patient which leads to the decision concerning the selection of the appropriate clinical pathway to be executed. The execution of the treatment scheme produces exceptions which are handled on real-time basis by the implemented software prototype.

More specifically, once the patient is admitted to the healthcare organization, its IT infrastructure should become aware of the data accompanying the specific patient. So, an initial data entry for the medical record dataset is performed. This procedure is performed either manually or automatically if the patient's medical record is received from another healthcare organization.

Once the clinical status of the patient is set, SEMPATH proposes an initial set of examinations to be performed. Afterwards, the result set of the test is inserted into the medical record of the patient. The system proposes an appropriate clinical pathway according to the diagnosis. So, the execution of the treatment scheme begins, under the constant inspection of SEMPATH prototype. Once an exception occurs, SEMPATH receives the exception information, executes the required rule-set and proposes the next "step" of the treatment scheme. The "step" derives from the following two categories:

- **Atomic process:** a single step procedure, executable by the process execution engine.
- **Complex process:** a multiple step procedure. It is a set of atomic processes without decision making. A complex process may contain parallel execution paths leading to a unified result.

The above-mentioned procedure is repeated during the execution of the treatment scheme of the patient. This way, the personalization of treatment for each patient is highly ensured, increasing the possibilities for the most suitable treatment.

After the implementation of the SEMPATH prototype, we are now in close cooperation with a University Hospital (LAIKO Hospital). We have modelled five (5) clinical pathways, defined the corresponding rules and we are in the process of real-life scenario execution.

5 CONCLUSIONS & FUTURE WORK

The existing practices and work performed in the area of clinical pathways has led to significant results. Nevertheless, clinical pathways are static procedures based on medical guidelines and organizational and operational knowledge. The current trend focuses on the adaptation of these static structures in case of exception occurrence.

The introduction of semantic technology provides further opportunities for clinical pathway adaptation. Semantics enable the representation of all required types of knowledge by the utilization of the corresponding ontologies. A significant amount of work has been done in the area of ontology creation. The developed ontologies cover several streams of medical knowledge.

However, this knowledge has to be combined with rules in order to handle the pathway exceptions. In this paper we presented SEMPATH prototype, a software platform based on semantics. SEMPATH introduces the creation of semantic rules in SWRL format which provide the basis for the rule-engine that handles the pathway exceptions. SWRL rules were created by the utilization of the SEMPATH ontology. The execution of the semantic rules

provides new knowledge objects that can be added to the existing ontology. Consequently, the knowledge enclosed in SEMPATH prototype is constantly updated and maintained by its routine operation.

Our intensions for further work can be presented in a three-fold structure:

- **Semantics Infrastructure:** our main aim concerning the evolution of SEMPATH semantics infrastructure is to proceed with the ontology enhancement. The enhancement will focus on organizational issues modelling and medical knowledge representation. Furthermore, our intention is to integrate existing medical ontologies so as to enrich our ontology model.
- **Pathway Modelling:** in the field of pathway modelling we plan to concentrate on simultaneous execution activities management and on providing different views of the pathway for each type of user. Moreover, our intention is to add semantic information in activities that will establish a priority weight model in order to perform more intelligent resource and activity management.
- **System evaluation and usability:** since the SEMPATH prototype is finalized and functional, we intend to perform real-life stress tests concerning its performance inside a healthcare organization. A real-life test will provide valuable results concerning the usability of the system, the performance and further enhancement of the implemented ontology and the further enhancement of the semantic rule-set.

ACKNOWLEDGEMENTS

This work has been partially funded by the European Commission with the IST 027065 contract (RIDE: A Roadmap for Interoperability of e-Health Systems in Support of COM 356 with Special Emphasis on Semantic Interoperability).

REFERENCES

Abidi, S., Chen, H., 2006. *Adaptable Personalized Care Planning via a Semantic Web Framework.*

ACPP Ontology, http://esw.w3.org/topic/HCLS/ACPPTaskForce

ACPP Task Force, http://esw.w3.org/topic/HCLS/ACPPTaskForce

Blaser, R., Schnabel, M., Biber, C., Baumlein, M., Heger, O., Beyer, M., Opitz, E., Lenz, R., Kuhn, K.A., 2007. *Improving pathway compliance and clinician performance by using information technology.* International Journal of Medical Informatics, Elsevier.

Campbell, H., Hotchkiss, R., Bradshaw, N., Porteous, M., 1998. *Integrated care pathways*. BMJ 316, 133-137.

Colaert, D., 2007. *Bringing the pieces together.*Towards semantic interoperability in e-Health Workshop, RIDE Project, Brussels.

Dadam, P., Reichert, M., Kuhn, K., 2000. *Clinical Workflows – the killer application for process-oriented information systems?*. Proceedings of 4th International Conference on Business Information Systems, Berlin, Springer, 36-59.

Greiner, Ul., Ramsch, J., Heller, B., Loffler, M., Muller, R., Rahm, Er., 2004. *Adaptive Guideline-based Treatment Workflows woth AdaptFlow.* Proceedings of the Symposium on Computerized Guidelines and Protocols, CGP 2004, Prague, IOS Press, 113-117.

Han, M., Thiery, T., Song, X., 2006. *Managing Exceptions in the Medical Workflow Systems.* ICSE '06, May 20-28, Shangai, China.

Kobayashi, M., Fussell, S. R., Xiao, Y., Seagull, F. J., 2005. *Work coordination, workflow, and workaround in a medical context.* Conference on Human Factors in Computing Systems.

Lenz, R., Blaser, R., Beyer, M., Heger, O., Biber, C., Baumlein, M., Schnabel, M., 2007. *IT support for clinical pathways – Lessons learned.* International Journal of Medical Informatics, Elsevier.

Lenz, R., Reichert, M., 2006. *IT Support for healthcare processes – premises, challenges, perspectives.* Data and Knowledge Engineering, ELSEVIER.

Miksch, S., Kosara, R., Seyfang, A., 2001. *Is Workflow Management Appropriate for Therapy Planning?*. Proceedings of EWGLP 2000, Amsterdam, IOS Press, 53-69.

OWL, http://www.w3.org/TR/owl-features/

Reichert M., Rinderle S., Dadam P., 2003. *ADEPT workflow management system: flexible support for enterprise-wide business processes*. Proc. Int'l. Conf. on Business Process Management (BPM'03), pp. 370–379.

Song, X., Hwong, B., Matos, G., Rudorfer, Ar., Nelson, C., Han, M., Girenkov, An., 2006. *Understanding Requirements for Computer-Aided Healthcare Workflows: Experiences and Challenges.* ICSE '06, May 20-28, Shanghai, China.

SWRL, http://www.w3.org/Submission/SWRL/

SWRLTab, http://protege.cim3.net/cgi/bin/wiki.pl?SWRLTab

Tucker, A. L., and Edmondson, A., 2002. *Managing Routine Exceptions: A Model of Nurse Problem Solving Behavior.* Advances in Health Care Management, 3, 87-113.

W3C, *Semantic Web Health Care and Life Sciences Interest Group,* http://www.w3.org/2001/sw/hcls/

THE STRUCTURED STORAGE OF ONCOLOGICAL CHEMOTERAPEUTIC REGIMENS

Contribution to Standardization of Therapeutic Procedures in Current Oncology

D. Klimes[1], L. Dusek[1], M. Kubasek[1], J. Novotny[2], J. Finek[3] and R. Vyzula[4]

[1] *Institute of Biostatistics and Analyzes, Masaryk University, Kamenice 126/3, Brno, Czech Republik*
klimes@iba.muni.cz, dusek@iba.muni.cz

[2] *Department of Oncology, General Teaching Hospital, Prague, Czech Republic*

[3] *Charles University Prague, Medical Faculty Pilsen, Pilsen, Czech Republic*

[4] *Masaryk Memorial Cancer Institute, Brno, Czech Republic*

Keywords: Cancer Chemotherapy Protocols, Antineoplastic Combined Chemotherapy Regimens, XML, Clinical Database.

Abstract: The aim of the chemotherapeutic regimens (CHR) digitalization project is the proposal of a universal structure and creation of a publicly accessible database of contemporary CHR as a universal utility for the communication and evaluation of contemporary and newly defined clinical schedules in anti-tumor chemotherapy. After analysis of contemporary anti tumor CHR a standard XML structure was proposed, which enables the recording of simple CHR from the field of chemotherapy in solid adult tumors, and also has the potential of recording the complex treatment protocols in the field of paediatric oncology. The resulting XML documents were saved on a web server. A publicly accessible CHR database was constructed. There were a total of 130 XML documents with definitions of individual CHR in the first phase. Linked to this data store, three examples of web applications were added to demonstrate the potential uses of this newly created database.

1 INTRODUCTION

Chemotherapy, along with surgery and radiotherapy is an irreplaceable part of the clinical treatment of oncological illnesses across nearly all diagnoses. To be more exact, we can define the term of chemotherapy as encountered in this article. It is the administration of individual preparations or combinations of preparations from the anatomical-therapeutic-chemical group (ATC) L01, which are anti-tumor preparations. This ATC group currently consists of nearly a hundred generic preparations, which are divided into 5 basic groups: Alkylating agents (L01A), Antimetabolites (L01B), Plant alkaloids and other natural products (L01C), Cytotoxic antibiotics and related substances (L01D) and Other antineoplastic agents (L01X). Whether these preparations are in practice applied either separately in monotherapy or in combination, we refer to them as a chemotherapeutic regimen (CHR). The more complicated plans for the application of cytostatics are classified in paediatric oncology as treatment protocols. Standard CHR are based on clinical studies and are published in the oncological journals (Goldberg et al., 2004) (Henderson et al., 2003) (Citron et al., 2003) and subsequently in national or international guidelines (NCCN, 2006) (CLS JEP, 2005).

However, CHR are not merely a list of applied cytostatics in administered doses; the definition of a CHR has its own basic rules, which are however not strictly defined anywhere. The main features of CHR include: doses of cytostatics are defined most often according to the surface area of the patient's body or their weight. CHR are applied in cycles, i.e. defined time segments of the treatment are repeated several times. There can be one or more repeated segments. The days of application are identified relative to the first day of each cycle (example Cycle 1- Day 1, Cycle 1- Day 8). These features are in practice routinely used (see Figure 1).

FEC chemotherapy
Cyclophosphamide 75 mg/m PO days 1-14
Epirubicin 60 mg/m IV days 1 & 8
5-Fluorouracil 500 mg/m IV days 1 & 8
With cotrimoxazole support.
Cycled every 28 days for 6 cycles.

Figure 1: An example of CHR clinical definition (NCCN Clinical Practice Guidelines in Oncology™ ©2006).

The aforementioned structure is frequently used, however not standardized, as to prevent a wider use of information technology. The vendors of a hospital information system can create a special application module for chemotherapy with a specific, internal data structure or chemotherapy data is stored only as a sequence of applications of cytostatics without specific details. This situation hinders on the one hand fast electronic transmission of new CHR and the general intercommunication between computer applications in this field, and on the other hand the more advanced use of technology, for instance in the field of the assessment and adherence to standard CHR in clinical practice. A possible solution is the creation of a structured data store of current CHR, which will mirror current clinical guidelines and which will be freely accessible to both clinics and HIS providers. This article describes the proposed structure of such a data store and the experience gained during its construction. The practical usage of such a database of CHR is demonstrated with the example of a web portal, whose components are web applications which use the CHR database as its datasource.

2 METHODS

Preliminary analysis of problems showed that the construction of a structured database of CHR calls for certain steps:

1) The proposal of the structure of CHR
2) The unique identification of standard CHR
3) The digitalisation of current standard CHR

CHR can be considered as a type of structured document, which is why the XML language was chosen for its recording. Standard XML offers a workable computational structure. In addition, it also provides tools for the internal validation of structures (XML schema) and tools for transforming documents into different, more user-friendly formats, for example HTML pages (XSLT). The proposed XML structure of elements and attributes was developed dynamically through the analysis of clinical definitions of standard CHR for individual oncological diagnoses. The source of the definitions was the National and International oncological guidelines (NCCN, 2006) (CLS JEP, 2005) and the internal source of Masaryk memorial Cancer Institut, a specialised hospital for the treatment of oncological diseases in the Czech Republic. The template for CHR was modelled using XML schemes, which enabled the definition of individual elements and attributes. The mandatory/optional properties and frequency of repetition of elements are defined by the maxOccurs and minOccurs indicators.

2.1 Header of CHR

The created XML structure of CHR consists of two parts. The first deals with the identification of CHR and the possibility of its use in oncological diagnosis (header). An example of this header is shown in figure 2.

```
<name>AC(Fisher)</name>
<sysname>(1;60.0;mg/m2;iv)A+(1;600.0;mg/m2;i
v)C&amp;amp;21
</sysname>
<diagnosis>
        <ICD10>C50</ICD10>
        <line>1</line>
        <purpose>adjuvant</purpose>
</diagnosis>
```

Figure 2: An example of CHR header definition.

The header contains the element name which was used for the clinical identification of CHR. This name was adopted either directly from clinical guidelines or from clinical identification in the information system of MMCI. These clinical names do not guarantee uniqueness and do not adhere to any strict rules. On the other hand the element sysname was added to the individual definitions of CHR on the basis of their internal structure. The detailed principles of their creation are described below.

The diagnosis element lists the individual oncological diagnoses that each individual CHR can treat. This is a complex element, which contains the following nested elements. The code for diagnoses is introduced according to the international classification ICD-10 in same named nested elements. The element line defines the CHR which is suitable (or approved) for specific lines of treatment.

The element purpose specifies whether a CHR is designed for adjuvant or palliative treatment. The elements line and purpose can, within the one element diagnosis, occur repeatedly in cases where the CHR is intended for more lines or for both basic purposes of treatment. In cases where the CHR is used for more diagnoses, the whole complex element Diagnosis is repeated.

2.2 Body of CHR

The second part of the structure describes the administration of given CHR (body). A standard CHR was divided into the following components

- The name of administered cytostatics
- Dosage of individual cytostatics
- Units of doses
- Method of administration
- Day relative to cycle of administration of cytostatic
- Duration of one cycle in days
- Total of completed cycles

Primarily proposed scheme for the body of the CHR is presented in figure 3.

```
<group>
<id_group>1</id_group>
<interval>21</interval>
<noc>4</noc>
<drug>
      .......
<drug>
<drug>
      .......
<drug>
</group>
<group>
<id_group>2</id_group>
      …….
</group>
```

Figure 3: The main frame of CHR body definition.

The element *interval* indicates the duration of one cycle of chemotherapy in days. The length of one cycle can be defined as the number of days between day D1 on one cycle and D1 on the following cycle. The actual length of the last cycle can only be determined from the last defined day of administration and cannot be compared with the length of the preceding cycles.

The element *noc* (number of cycles) shows the total of applied cycles. This parameter is limited by clinical guidelines to a small number of CHR. These guidelines often merely provide a recommendation for the repetion of cycles. In practice the number of applied cycles is decided by the actual state of health of the patient. In cases where details were not explicitly known, the value of this element was set to 0.

The complex element *drug* describes the application of individual cytostatics within the framework of one cycle of chemotherapy. Since many cytostatics are applied in CHR, the element *drug* is referred to as a recurrent element. The element *drug* encapsulates the nested elements for labelling cytostatics, their dosage, day of administration and method of administration. As soon as we try to identify individually applied cytostatics, the problem of their individual classification arises. Existing practice is to cite the full generic name of the cytostatic, which in certain cases involves the brand name of the medication. The ATC codes of identification of cytostatics are not used in clinical practice, however they are ideal for computer processing. We therefore decided to include in the XML structure both an element for the generic name of cytostatics (*name*) and an element for the ATC code (*ATC).* The element *name* can be inserted for each cytostatic repeatedly with the attribute *lang* for various language versions of classification. For the purpose of the systematic naming of CHR (described below) the element *abbr* was also defined for the abbreviated names of cytostatics.

Doses of cytostatics, units of dosage of cytostatics, method of administration and relative day of administration were included within the complex element *administration*, whose title was shortened to *adm* for practical reasons. The element *dose* was defined for the dosage of cytostatics as a real number, for units of dosage the enumerative element *unit,* for the method of administering the enumerative element *mode* and for the day of administering the whole number element *unit,* respectively the two whole number elements *start_day* and *end_day.*

In anti tumour chemotherapy, the dosage of cytostatics is most commonly defined by the calculation of the surface area of the patient or by their weight. For this reason the element *unit* was defined with the values mg/m^2 and mg/kg. Carboplatin has special dosage, where the dose is defined in AUC. The resulting dose of this drug is calculated depending on the laboratory value parameters creatinine clearance (CrCl) according to the formula:

dose(mg) = target AUC (mg/ml/min) * (CrCl +25) (ml/min)

Two basic methods of administering cytostatics (modes) are used, per oral and intravenous administration. More detailed categorization can be considered, for instance distinguishing between intravenous administrations according to the length of infusion. In the basic proposal there is the difference between bolus administration and infusion. For example in the case of the regimens FOLFOX 4 and FOLFOX 6, which are used in the treatment of colorectal carcinoma, it is necessary to differentiate between the bolus dosage of fluorouracil from its subsequent longterm infusion. The element *mode* can thus include the values iv, iv-bolus and po.

Day of administration is in clinical practice presented as Dx, where x is the consecutive number of days from the administration of the first preparation in a cycle. Individual cytostatics can be repeatedly administered within a cycle, either on chosen days (e.g D1, D8) or daily in the course of an appointed time period (e.g D1-D14). For the first variant, the element *day* can be repeated within the complex element *adm,* and in the second variant there are, in place of the element day, two elements *start_day* and *end_day.* An example of the complex element *drug* is illustrated in figure 4.

```
<drug>
<name lang="cz">Cyklofosfamid
  </name>
  <name lang="eng">Cyclophosphamide
  </name>
   <atc>L01AA01</atc>
  <abbr>C</abbr>
   <adm>
        <day>1</day>
        <dose>600</dose>
        <unit>mg/m2</unit>
        <mode>iv</mode>
   </adm>
</drug>
```

Figure 4: An example of the complex element *drug.*

For ‘multigroup’ regimens such AC+paclitaxel (see figure 5) it was necessary to extend the presented concept.

Doxorubicin 60 mg/m IV day 1
Cyclophosphamide 600 mg/m IV day 1
Cycled every 21 days for 4 cycles.
Followed by
Paclitaxel 175-225 mg/m by 3 h IV infusion day 1
Cycled every 21 days for 4 cycles.

Figure 5: An example of multigroup CHR (NCCN Clinical Practice Guidelines in Oncology™ ©2006).

For this type of CHR another encapsulating structure was added to the XML scheme. This took the form of the specific complex element *group,* which contains all the defined elements of the body of the CHR and can be repeated. For the identification of groups, the element *id_group* was added which contains the consecutive number of the group. For multigroup CHR it is necessary that the element *noc* must not be zero, at least for each group except the last.

2.3 Systematic Naming of CHR

To prevent duplication in the database of CHR, a concept was sought after, which would ensure the individual identification of each of the stored CHR. The identification of CHR used in clinical practice seemed unsuitable, because as often happens, one CHR has more than one name, or one name refers to more than one CHR.

A unique standard naming of CHR was inspired by Logical Observation Identifiers Names and Codes (LOINC). LOINC is a structured classification of laboratory methods (Huff et al., 1998). There is a systematic name for each item (in this example the laboratory method) consisting of individual components which the method uniquely refers to. In the case of LOINC, the names of laboratory methods include the components, property, timing, system precision and method. Similarly, it is possible to create a unique identification system of CHR. The following requirements were necessary to be taken into account within the proposals for the rules for the systematic creation of naming of CHR:

- The naming has to be unique
- The naming must be automatically generated from definitions of the CHR
- All key components of the scheme must be coded into the name
- The name must remain "human readable"

The schematic name is created according to the following syntactic rules:

- Administered drugs are classified with a unique abbreviation
- Drugs are alphabetically sequenced according to abbreviations and are divided by the symbol plus (+)
- After the listing of the drug the duration of cycle in days is added after the symbol(&)
- For every drug, the following items are defined in round brackets separated by a semicolon (;)

The first entry in the brackets is the day of administration. It can be in the form of a number (1), a list of numbers separated by commas (1, 8) or an interval (1-14). The second entry is the dosage of the medication. The third entry is the abbreviated name for the method of administration. The fourth entry is the unit of dosage.

Abbreviations used for cytostatic medications are summarized in Table 1.

Abbreviations used are parts of the proposal for the standardization of structures of CHR, because currently the standard for the abbreviated identification of cytostatics has not been found (on the webpages of NCI (NCI Drug Dictionary National Cancer Institute, 2005) only recommended abbreviations can be found and all the well known synonyms for given preparations). Clinically established abbreviations are often diagnosis-specific, for example cisplatin is listed under the letter P in CHR such as BIP or BEP, in the regimen M-VAC it is classified under the letter C, which is however in the majority of cases used as the abbreviation for cyclophosphamide. Due to the fact that the number of well known cytostatics is very similar to the number of existing chemical elements for which two symbols are sufficient for the abbreviated symbols, a similar concept was used for the identification of cytostatics. The names of the most frequently used cytostatics are written in the table, each is recorded with a NCI abbreviation, ATC code and proposed two letter identification, which issues from the generic name of the cytostatic.

For multigroup CHR the concept was further developed with square brackets enclosing individual groups. The number of cycles is indicated before the brackets separated by asterisks (*). Groups are separated with the symbol +. An example of the identification of the CHR AC+paclitaxel is illustrated in figure 6.

Table 1: Summary of most frequently used cytostatic agents.

Cytostatic agent	ATC - code	NCI abbreviation*	Used abbreviation
bevacizumab	L01XC07	?	Be
bleomycin	L01DC01	BLEO	B
busulfan	L01AB01	BU, BUS	Bu
capecitabine	L01BC06	CAPE	Ca
carboplatin	L01XA02	CBDCA	Cb
carmustine	L01AD01	BCNU	Bc
cetuximab	L01XC06	MOAB C225	Ce
chlorambucil	L01AA02	CHL, CLB	Cl
cisplatin	L01XA01	CDDP	P
cyclophosphamide	L01AA01	CTX	C
cytarabine	L01BC01	ARA-C	Cy
dactinomycin	L01DA01	DACT	Ac
dacarbazine	L01AX04	DTIC	Dc
daunorubicin	L01DB02	DNR	Dn
docetaxel	L01CD02	TXT	Dt
doxorubicin	L01DB01	DOX	A
epirubicin	L01DB03	EPI	E
erlotinib	L01XX34	OSI 774	Er
estramustine	L01XX11	EM	Em
etoposide	L01CB01	VP-16	Et
fludarabine	L01BB05	FAMP	Fl
fluoruracil	L01BC02	5-FU	F
gefitinib	L01XX31	ZD 1839	Ge
gemcitabin	L01BC05	dFdC	G
ifosfamide	L01AA06	IFF, IFO	If
irinotecan	L01XX19	CPT-11	I
melphalan	L01AA03	L-PAM	Ml
methotrexate	L01BA01	MTX	M
mitomycin	L01DC03	MITO	Mi
mitoxantrone	L01DB07	DHAD	Mx
oxaliplatin	L01XA03	1-OHP, L-OHP	Oh
paclitaxel	L01CD01	TAX	Ta
pemetrexed	L01BA04	LY231514	Pe
prednimustine	L01AA08	?	Pr
procarbazine	L01XB01	PCB	Pc
raltitrexed	L01BA03	?	Ra
rituximab	L01XC02	MOAB IDEC-C2B8	Ri
temozolomide	L01AX03	TMZ	Tm
thiotepa	L01AC01	TSPA	Ts
topotecan	L01XX17	TOPO	To
trastuzumab	L01XC03	MOAB HER2	Tr
vinblastine	L01CA01	VBL	V
vincristine	L01CA02	VCR	Vc
vinorelbine	L01CA04	VNB	Vn

* NCI Drug Dictionary (NCI Drug Dictionary National Cancer Institute, 2005)

4*[(1;60.0;mg/m2;iv)A+(1;600.0;mg/m2;iv)C&21]+
4*[(1;175.0;mg/m2;iv)Pt&;21]

Figure 6: An example of the identification of the CHR AC+ paclitaxel

This identification of CHR is stored in a XML file in the element *sysname.* The element *sysname* acts as a unique identifier, a type of "fingerprint" of

the CHR and its primary function is to prevent duplication in the data system.

3 RESULTS

A total of 160 CHR were entered into the database. The definitions were formed according to the Czech national guidelines for the cytostatic treatment of solid tumors (CLS JEP, 2005). The validity of entries according to International standards was verified with reference to the International guidelines NCCN (NCCN, 2006).

3.1 CHR Library Applications

As a demonstration of the use of structured records of definitions of CHR a publicly accessible web application was developed with three basic functions: The Central Library of Chemotherapeutic Regimens, Dose Intensity (DI) Calculator and Therapy Organiser.

3.1.1 Search Engine

The Central Library is a simple search engine, that, according to user entered criteria, searches and displays the definition of the corresponding regimen. Registered users have the possibility to add textual commentaries to each CHR with supplementary information, while for non registered users all information is presented in a read only format.

3.1.2 Dose-Intensity Calculator

The DI Calculator enables users to calculate the dose-intensity for selected CHR according to the methods in (Hryniuk et al., 1984), and to compare this with the actual intensities of cytostatics administered to the patient in question.

3.1.3 Therapy Organizer

The Therapy Organizer enables users to devise time plans for the administration of chosen CHR. It is possible to display and print this plan in the form of a calendar with suggested days and dosages for individual cytostatics. The functions mentioned are interconnected, for example, search results from the Central Library of CHR can be directly used to create a time plan in Therapy Organizer.

The applications are accessible at the internet address
http://dios.registry.cz/?sec=software&lang=en

3.2 CHR Derivation

Thanks to a structured CHR library it is possible to derive standard regimen only from a list of applicated drugs and dates of administration. This is useful when dose-intensity is evaluated and available data doesn't contain name of standard regimen. During pilot tests there was success in correctly deriving 98% of initial CHR from 180 patients who had been administered chemotherapy for breast carcinoma.

4 CONCLUSIONS

CHR and their administration are routine practice in contemporary oncology. The development of a structured, electronic database of standard CHR can help the faster propagation of information about new CHR and at the same time enable assessment of their adherence in clinical practice. The database is created from XML documents, where every file represents one CHR.
Unlike other printed or electronic sources about CHR, this database contains only clear, structured records of regimes. These records are inserted in cooperation with expert oncologists. The result is a new, always up-to-date information source that forms the base for Dose Intensity Analysis and also can be used in other computer applications in anti-tumour therapy area.

ACKNOWLEDGEMENTS

Project DIOS, which addresses problems in anti tumour chemotherapy, is supported by the grant 2608 from Ministry of Education of the Czech Republic and by the Amgen Inc.

REFERENCES

Goldberg, R. M., D. J. Sargent, R. F. Morton, C. S. Fuchs, R. K. Ramanathan, S. K. Williamson, B. P. Findlay, H. C. Pitot, S. R. Alberts, 2004. A Randomized controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer, Journal of Clinical Oncology 22, no. 1: 23-30.

Henderson, I. C., D. A. Berry, G. D. Demetri, C. T. Cirrincione, L. J. Goldstein, S. Martino, J. N. Ingle, M. R. Cooper, D. F. Hayes, K. H. Tkaczuk, G. Fleming, J. F. Holland, D. B. Duggan, J. T. Carpenter, E. Frei, R.

L. Schilsky, W. C. Wood, H. B. Muss, L. Norton, 2003. Improved outcomes from adding sequential paclitaxel but not from escalating doxorubicin dose in an adjuvant chemotherapy regimen for patients with node-positive primary breast cancer, Journal of Clinical Oncology 21, no. 6 (2003): 976-83.

Citron, M. L., D. A. Berry, C. Cirrincione, C. Hudis, E. P. Winer, W. J. Gradishar, N. E. Davidson, S. Martino, R. Livingston, J. N. Ingle, E. A. Perez, J. Carpenter, D. Hurd, J. F. Holland, B. L. Smith, C. I. Sartor, E. H. Leung, J. Abrams, R. L. Schilsky, H. B. Muss, L. Norton, 2003. Randomized trial of dose-dense versus conventionally scheduled and sequential versus concurrent combination chemotherapy as postoperative adjuvant treatment of node-positive primary breast cancer: First report of intergroup trial C9741/cancer and leukemia group B trial 9741, Journal of Clinical Oncology 21, no. 8,p. 1431-39.

NCCN Clinical Practice Guidelines in Oncology™ NCCN, (accessed 1.10.2006). Available from http://www.nccn.org/professionals/physician_gls/f_guidelines.asp.

Zasady cytostaticke lecby malignich onkologickych onemocneni: Ceska onkologicka spolecnost CLS JEP, 2005.

Huff, S. M., R. A. Rocha, C. J. McDonald, G. J. De Moor, T. Fiers, W. D. Bidgood, Jr., A. W. Forrey, W. G. Francis, W. R. Tracy, D. Leavelle, F. Stalling, B. Griffin, P. Maloney, D. Leland, L. Charles, K. Hutchins, J. Baenziger, 1998. Development of the Logical Observation Identifier Names and Codes (LOINC) vocabulary, Journal of the American Medical Informatics Association 5, no. 3, p. 276-92.

NCI Drug Dictionary National Cancer Institute, 2005 (accessed 12.9. 2005). Available from http://www.cancer.gov/drugdictionary/.

Hryniuk, W., H. Bush, 1984. The Importance of Dose Intensity in Chemotherapy of Metastatic Breast-Cancer, Journal of Clinical Oncology 2, no. 11, p. 1281-88.

MINING OF HEALTH INFORMATION FROM ONTOLOGIES

Maja Hadzic
Research Lab for Digital Health Ecosystems, Curtin Universityof Technology, Austria
m.hadzic@curtin.edu.au

Fedja Hadzic, Tharam Dillon
Digital Eciossytems and Business Intelligence Institute (DEBII), Curtin Universityof Technology, Austria
fedja.hadzic@postgrad.curtin.edu.au, t.dillon@curtin.edu.au

Keywords: Data mining, Tree mining, Ontology, Ontology mining, Human disease ontology, Human disease study, Health information system.

Abstract: Data mining techniques can be used to efficiently analyze semi-structured data. Semi-structured data are predominantly used within the health domain as they enable meaningful representations of the health information. Tree mining algorithms can efficiently extract frequent substructures from semi-structured knowledge representations. In this paper, we demonstrate application of the tree mining algorithms on the health information. We illustrate this on an example of Human Disease Ontology (HDO) which represents information about diseases in 4 'dimensions': (1) disease types, (2) phenotype (observable characteristics of an organism) or symptoms (3) causes related to the disease, namely genetic causes, environmental causes or micro-organisms, and (4) treatments available for the disease. The extracted data patterns can provide useful information to help in disease prevention, and assist in delivery of effective and efficient health services.

1 INTRODUCTION

New modern techniques are providing huge, rapidly accumulating amounts of information. To extract and analyze the data poses a much bigger challenge for researchers than to generate the data (Holloway *et al.*, 2002). Experienced scientists and doctors are overwhelmed with this situation.

There is a need for an intelligent and efficient system to make use of all the available information. The true value of this information can be significantly increased through smart information processing and analysis. Such systems could play a crucial role in filtering the flood of data to the point where human experts could apply their knowledge sensibly.

Information technologies must be effectively implemented within health domain. In their paper, Horvitz-Lennon *et al*. (2006) state that we need to fully embrace information technology and its potential for improving service efficiency and develop a better information infrastructure for the patient's care.

Data mining is a set of processes that is based on automated searching for actionable knowledge buried within a huge body of data. Data mining techniques extract information and find hidden patterns and behaviors, and support making predictive models for decision making and new discoveries. Within the biomedical and health field, data mining techniques have been predominately used for tasks such as text mining, gene expression analysis, drug design, genomics and proteomics (Zaki *et al.*, 2003). The data analysis necessary for microarrays has necessitated data mining (Piatetsky-Shapiro & Tamayo, 2003). Recently, use of data mining methods has been proposed for the purpose of mapping and identification of complex disease loci (Onkamo & Toivonen, 2006). However, the proposed methods are yet to be implemented. Frequent pattern analysis has been a focused theme of study in data mining. A lot of algorithms and methods have been developed for mining frequent sequential and structural patterns (Han & Kamber, 2001; Agrawal & Srikant, 1994; Tan *et al.*, 2006a). Implementation of data mining techniques within health domain could help in disease prevention and

assist in delivery of effective and efficient health services.

Within the data mining field, tree mining has recently attracted lots of interest. Our work in the tree mining field is characterized by a Tree Model Guided (TMG) (Tan *et al.*, 2005; Tan *et al.*, 2006b) candidate generation approach. This non-redundant systematic enumeration model uses the underlying tree structure of the data to generate only valid candidates. Using the general TMG framework a number of algorithms were developed as follows. MB3-Miner (Tan *et al.*, 2005) mines ordered embedded subtrees while IMB3-Miner (Tan *et al.*, 2006b) can mine both, induced or embedded ordered subtrees by using the level of embedding constraint. Razor algorithm (Tan *et al.*, 2006c) was developed for mining of embedded subtrees where the distances of nodes relative to the root of the original tree need to be considered. UNI3 algorithm (Hadzic *et al.*, 2007a) mines induced unordered subtrees. Our algorithms were applied on large and complex tree structures and their scalability was experimentally demonstrated (Tan *et al.*, 2005; Tan *et al.*, 2006b). From the application perspective, in (Hadzic *et al.*, 2006) we have applied our tree mining algorithm for the analysis of Protein Ontology (Sidhu *et al.*, 2004) database for Human Prion proteins which was represented in XML format. In this paper, we explain how tree mining techniques can be applied within the health domain for deriving useful knowledge patterns that can help disease prevention and management.

2 HUMAN DISEASE ONTOLOGY

We designed Human Disease Ontology (GHDO) (Hadzic & Chang, 2005) to have the following four branches or subontologies:

- *disease types*, describing different types of a disease;
- *phenotype*, describing disease symptoms;
- *causes* responsible for that disease that can be genetic, environmental and/or microorganism;
- *treatments*, providing an overview of all treatments possible for a particular disease;

Top-level hierarchy of the HDO is illustrated in Figure 1.

The information presented in Figure 1 state that a disease may have different types, and associated subtypes and sub-subtypes. For each disease, there is a corresponding phenotype (or observable characteristics of an ill individual), namely symptoms of a disease. Disease cause can be genetic (genotype), environmental or a microorganism. Genetic causes can be a mutated gene, a complex of genes or DNA region of interest. DNA region of interest is a region in the DNA sequence that potentially contains a gene responsible for the disease. This region needs to be further examined in order to correctly locate the mutated gene. Environmental causes of a disease can be stress, climate, drugs misuse, family conditions or economic condition. Microorganisms that may cause a disease may be virus or bacteria. Possible treatments for a disease can be drug therapy, chemotherapy, surgery, psychotherapy or physiotherapy.

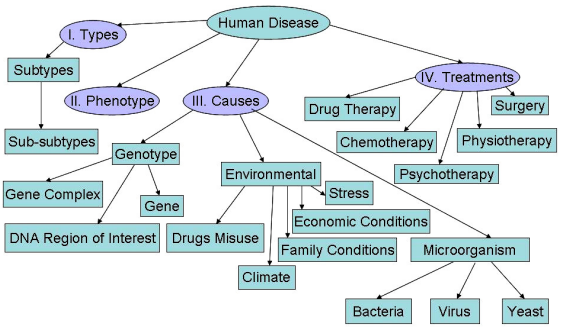


Figure 1: Top-level hierarchy of the HDO.

Researchers in the medical ontology-design field have developed different terminologies and ontologies in many different areas of medical domain. In order to obtain some uniformity across different ontologies, definitions from other published and consensual ontologies can be reused (Noy & Musen, 2000). We can use other ontologies such as LinkBase (Montyne, 2001) and UMLS (Bodenreider, 2004). These ontologies contain over million concepts, and we do not require all of them. The four different branches (subontologies) of the GHDO ontology can serve as a reference point against which the concepts from the existing ontologies can be reorganized, aligned and merged. Bimolecular ontologies such as TAMBIS ontology (Stevens *et al.*, 2002) that represents general knowledge in regard to proteins and associated genes may be suitable to cover the ontology part in regard to genetical disease causes. So, we can use terminology from existing ontologies but select and organize the concepts in a way that can be used in our application.

Note that the HDO has a tree (hierarchy) structure which allows it to be analyzed using some available tree mining techniques. The current ontology languages allow the use of graph structures to represent the domain knowledge in an ontology. A large portion of current ontologies have predominantly hierarchical structures. Furthermore, it is often the case that the graph-structured knowledge representation can be modelled using tree structures without losing too much semantics. The root of the complexity of mining graph structures is in the existence of cycles, and in many cases the number of cycles in graph instances is limited. The complexity of processing tree structures tends to be more manageable and is one of the promising directions towards automatic analysis of ontologies.

We assume that the available health information can be represented according to the four HDO 'dimensions' or subontologies. In the rest of our paper, we will use the term 'HDO instance' to refer to a specific record found within a given database/application that can be represented using the HDO structure. In our previous publications such as (Hadzic & Chang, 2005), this 'HDO instance' corresponds to Specific Human Disease Ontology.

3 TREE MINING CONCEPTS

The aim of this section is to provide the definitions of some basic tree concepts necessary for understanding the current work. Please refer to (Tan *et al.*, 2005; Tan *et al.*, 2006b; Hadzic *et al.*, 2007a) for a more extensive overview of the tree mining area including the discussion of implementation issues and algorithm comparisons. A tree is a special type of graph where no cycles are allowed. It consists of a set of *nodes (or* vertices) that are connected by *edges*. Each edge has two nodes associated with it. A *path* is defined as a finite sequence of edges and in a tree there is a single unique path between any two nodes. The *length of a path p* is the number of edges in *p*. A rooted tree has its top-most node defined as the *root* that has no incoming edges and for every other node there is a path between the root and that node. A node *u* is said to be a *parent* of node *v*, if there is a directed edge from *u* to *v*. Node *v* is then said to be a *child* of node *u*. Nodes with no children are referred to as *leaf* nodes and otherwise they are called *internal nodes*. If for each internal node, all the children are ordered, then the tree is an ordered tree. The problem of frequent subtree mining can be generally stated as: Given a tree database T_{db} and minimum support threshold (σ), find all subtrees that occur at least σ times in T_{db}.

A HDO instance can be captured by an OWL document analogous to the one shown in Figure 2. In our example, the actual information content can be viewed as an ordered labeled tree. This specific instance aims to capture the information about causal factors of a human disease and corresponds with the Causes subontology shown in Figure 1.

```
<?xml version="1.0" ?>
- <rdf:RDF xmlns="http://www.owl-ontologies.com/cause
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#" >
    xmlns:p1="http://www.owl-ontologies.com/assert.ow
    <owl:Ontology rdf:about="" />
  - <rdfs:Class rdf:ID="Virus">
    - <rdfs:subClassOf>
        <rdfs:Class rdf:ID="Microorganism" />
      </rdfs:subClassOf>
    </rdfs:Class>
  - <rdfs:Class rdf:ID="Genotype">
    - <rdfs:subClassOf>
        <rdfs:Class rdf:ID="Cause" />
      </rdfs:subClassOf>
    </rdfs:Class>
  - <rdfs:Class rdf:about="#Microorganism">
      <rdfs:subClassOf rdf:resource="#Cause" />
    </rdfs:Class>
  - <rdfs:Class rdf:ID="Climate">
    - <rdfs:subClassOf>
        <rdfs:Class rdf:ID="Environmental" />
      </rdfs:subClassOf>
    </rdfs:Class>
```

Figure 2: OWL representation of part of the HDO subontology i.e. Causes subontology.

A tree can be denoted as *T(V, L, E)*, where:
(1) *V* is the set of *vertices* or *nodes*;
(2) *L* is the set of *labels* of vertices, for any vertex $v \in V$, *L(v)* denotes the label of *v*; and
(3) $E = \{(x,y) | x,y \in V\}$ is the set of edges in the tree.

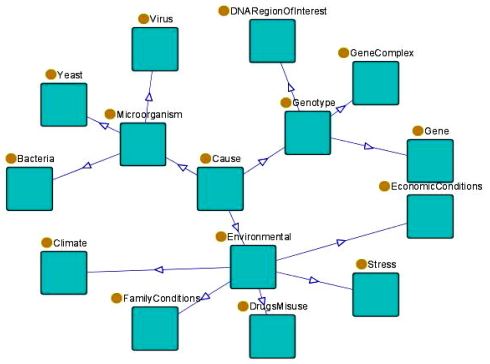


Figure 3: Causes subontology of HDO (viewed in Protégé).

In Figure 3, we represent the OWL document from Figure 2 as a tree with "Cause" as the root node.

Most of the current tree mining algorithms are mainly focused on extracting induced and embedded subtrees. An induced subtree preserves the parent-child relationships of each node in the original tree. In addition to this, an embedded subtree allows a parent in the subtree to be an ancestor in the original tree and hence ancestor-descendant relationships are preserved over several levels. Formal definitions follow:

A tree $T'(V', L', E')$ is an **induced** subtree of a tree $T(V, L, E)$ iff (1) $V' \subseteq V$, (2) $E' \subseteq E$, (3) $L' \subseteq L$ and $L'(v)=L(v)$, (4) $\forall v' \in V'$, $\forall v \in V$, v' is not the root node, and v' has a parent in T', then parent(v')=parent(v).

A tree $T'(V', L', E')$ is an **embedded** subtree of a tree $T(V, L, E)$ iff (1) $V' \subseteq V$, (2) if $(v_1,v_2) \in E'$ then $parent(v_2) = v_1$ in T', only if v_1 is ancestor of v_2 in T and (3) $L' \subseteq L$ and $L'(v)=L(v)$.

The subtrees can further be distinguished depending on the order of sibling nodes. An **ordered** subtree preserves the left-to-right ordering among the sibling nodes in the original tree. In an **unordered** subtree the left-to-right ordering among the sibling nodes does not need to be preserved. The order of the sibling nodes (and the subtrees rooted at sibling nodes) can be exchanged and the resulting subtree would be considered the same. Examples of different subtree types related to the Figure 3 are given in Figure 4.

The available support definitions are transaction based, occurrence-match, and hybrid support (Tan *et al.*, 2006b; Hadzic *et al.*, 2007b). **Transaction based support** only checks for the existence of items in a transaction while **occurrence-match** support takes the repetition into account and counts the item occurrences in the database as a whole. Using **hybrid support** with a threshold set to $x|y$, a subtree will be considered as frequent iff it occurs at least y times in x number of transactions.

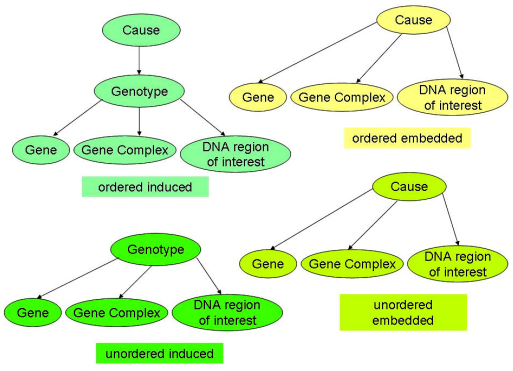


Figure 4: Example of different subtree types.

4 MINING OF HEALTH INFORMATION

In order to apply tree mining to the problem, the following five phases take place: data selection and cleaning, data formatting, tree mining, pattern discovery, and knowledge testing and evaluation.

In the first phase, we focus on the Data Selection and Cleaning. Most of the databases also contain information that is not needed by the application. The irrelevant information, as well as noise and inconsistent data, should be removed from the data set to be mined. Extra care should be taken during this process since some data may appear to be noisy but in fact represents a true exceptional case.

In the second phase, Data Formatting, all the collected data needs to be represented using the same format that is understandable by the tree mining algorithm used.

The third phase is concerned with applying the tree mining algorithms for extracting interesting patterns from now a clean and correctly formatted dataset. One needs to consider what particular type of subtree is the most suitable one to mine for satisfying the application needs. When the data to be mined comes from one organization, the format and ordering of the presented information is expected to be the same, and hence mining of ordered subtrees will be sufficient. On the other hand, if the collected data originate from separate organizations, then mining of unordered trees would be more suitable. The organizations could order their concepts differently, but since an unordered subtree with different order among the sibling nodes is still considered as the same candidate, the common characteristics of a particular illness would still be found. Another choice to be made is whether induced or embedded subtrees should be mined. When mining health related data it is important that particular information stays in the context where it occurred. With respect to tree patterns this implies that the relationship of nodes in the extracted subtrees should be limited to parent-child relationships. Allowing ancestor-descendant relationships would result in information loss about the context where a particular disease characteristic occurred. Some attributes in the dataset may have a similar set of values and by mining induced subtrees there the attribute to which each value belongs will be indicated in the extracted pattern.

As a final consideration the support definition chosen should be dependant on the way that the data is organized. Next we provide three common ways

in which the data could be presented, and indicate the suitable support definition that should be used.
Case 1: Each HDO instance is stored as a separate subtree in the OWL document and HDO instances describing different diseases are stored in separate documents. In this case both, occurrence match or transaction based support would be suitable.
Case 2: Each HDO instance is stored as a separate subtree in the OWL document but now one OWL document contains all HDO instances for all investigated diseases. Here the transactional support would be more appropriate.
Case 3: A collection of HDO instances related to one particular disease is always contained in a separate subtree of an OWL document. Hybrid support definition is most suitable in this case.
In cases 1 and 2, the minimum support threshold should be close to the number of HDO instances that the dataset contains about a particular disease, in order to find the commonality among all the records related to that particular disease. However, since noise is often present in the data, the support can lowered but not too much so that irrelevant factors are not picked up as important. In case 3, the number of diseases described would be used as the transactional part of the hybrid support, while the minimum occurrence of a subtree within each transaction should reflect the number of HDO instances.

As a common data mining practice, the data set at hand could be split into two subsets. One is used for deriving the knowledge model ('internal data' from Figure 5) and while the second one is used for testing the derived knowledge model ('external data' from Figure 5). When possible the data collected by another organization can be used as external data. During Pattern Discovery phase, new knowledge about specific disease(s) emerges. For example, these results may help to associate precise combinations of genetic and environmental factors with a specific disease type. The results could increase the understanding of the disease under study and make a breakthrough in the research, control and prevention of this disease. Knowledge Testing and Evaluation is illustrated in Figure 5. The 'external data' is used to verify a formed hypothesis before it can extend the current body of knowledge. The choice of the tree mining parameters often affects the nature and granularity of the obtained results. In cases where the hypothesis is not supported by the 'external data', the parameters will be adjusted and the previous steps alternated.

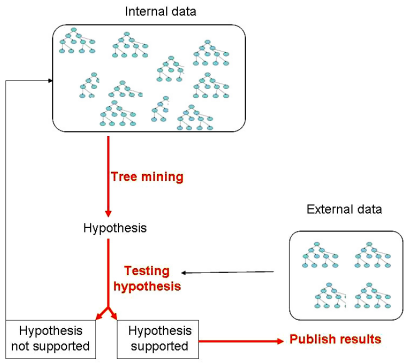


Figure 5: Testing and evaluation of the derived knowledge.

5 CONCLUSIONS

Data mining systems in general could play a crucial role in deriving knowledge and assisting in the prevention, diagnosis, treatments and control of human diseases. In this paper, we have illustrated application of the data mining technology within the health domain. We have explained how the tree mining algorithms can be effectively used for mining of information about human diseases. We aim to apply those techniques on the real-word data.

The newly derived knowledge could help in prevention of human diseases and assist in delivery of effective and efficient health services. Various community groups would greatly benefit from application of data mining techniques on a large scale, e.g. *physician* would receive support in early diagnosis and treatment of diseases; *patients* would receive support in dealing with, managing and treating disease; accurate, reliable and up-to-date information would be available for *general public*, and *medical researchers* would receive support in advancing their research. Furthermore, the *cost* of the health budget would be significantly reduced by providing better information use (Garber, 2006). Such systems go some way to delivering what Patel *et al* (2006) say is necessary to transform the quality of health care. They improve the infrastructure for evidence-based interventions and provide innovation for improvement in health care.

REFERENCES

Agrawal, R, Srikant, R., 1994. Fast algorithms for mining association rules. *Very Large Databases*, Chile.

Bodenreider, O., 2004. The Unified Medical language System (UMLS): integrating biomedical terminology, *Nucleic Acids Res*, vol. 32, no. 1, pp. 267-270.

Garber, A.M., 2006. PERSPECTIVE: To Use Technology Better. *Health Affairs*, pp. W51-W53.

Hadzic, F., Dillon, T.S., Sidhu, A., Chang, E., Tan, H., 2006. Mining Substructures in Protein Data. *ICDM'06 Wkshp on Data Mining in Bioinformatics*, China.

Hadzic, F., Tan, H., Dillon, T.S., 2007a. UNI3 – Efficient Algorithm for Mining Unordered Induced Subtrees Using TMG Candidate Generation. *IEEE Symposium on Computational Intelligence and Data Mining*, USA.

Hadzic, F., Tan, H., Dillon, T.S., Chang, E., 2007b. Implications of frequent subtree mining using hybrid support definition, *Data Mining & Information Engineering*, UK.

Hadzic, M., Chang, E., 2005. Ontology-based Support for Human Disease Study. *Proc. of the 38th Hawaii Int'l Conf. on System Sciences (HICSS-38)*, USA.

Han, J., Kamber, M., 2001. *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann.

Holloway, A., Van Laar, R.K., Tothill, R.W., and Bowtell D., 2002. Options available - from start to finish - for obtaining data from DNA microarrays II. *Nature Genetics Supplement*, vol. 32, pp. 481-489.

Horvitz-Lennon, M., Kilbourne, A.M., Pincus, H.A., 2006. From Silos To Bridges: Meeting The General Health Care Needs Of Adults With Severe Mental Illnesses. *Health Affairs*, vol. 25, no. 3, pp. 659-669.

Montyne, F., 2001. The importance of formal ontologies: a case study in occupational health. *Proc. of Int'l Wkshp on Open Enterprise Solutions: Systems, Experiences, and Organizations (OES-SEO2001)*, Italy.

Noy, N.F., Musen, M.A., 2000. PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment, *Proceedings of the seventeenth national conference on Artificial Intelligence (AAAI-2000)*, USA.

Onkamo, P., Toivonen, H., 2006. A survey of data mining methods for linkage disequilibrium mapping. *Human genomics*, vol. 2, no. 5, pp. 336-340.

Patel, K.K., Butler, B., Wells, K.B., 2006. What Is Necessary To Transform The Quality Of Mental Health Care. *Health Affairs*, vol. 25, no. 3, pp. 681-693.

Piatetsky-Shapiro, G., Tamayo, P., 2003. Microarray Data Mining: Facing the Challenges. *SIGKDD Explorations*, vol. 5, no. 2, pp. 1-6.

Sidhu, A.S., Dillon, T.S., Sidhu, B.S., Setiawan, H., 2004. A Unified Representation of Protein Structure Databases. *Biotechnological Approaches for Sustainable Development*, pp. 396-408.

Stevens, R., Baker, P., Bechhofer, S., Ng, G., Jacoby, A., Paton, N.W., Goble, C.A., Brass, A., 2002. TAMBIS: Transparent Access to Multiple Bioinformatics Information Sources. *Bioinformatics*, vol. 16, no. 2, pp.184-86.

Tan, H., Dillon, T.S., Hadzic, F., Chang, E., Feng, L., 2005. MB3-Miner: mining eMBedded sub-TREEs using Tree Model Guided candidate generation. *Workshop on Mining Complex Data*, USA.

Tan, H., Dillon, T.S., Hadzic, F., Chang, E., 2006a. SEQUEST: mining frequent subsequences using DMA Strips. *Data Mining and Information Engineering*, Czech Republic.

Tan, H., Dillon, T.S., Hadzic, F., Feng, L, Chang, E., 2006b. IMB3-Miner: Mining Induced/Embedded subtrees by constraining the level of embedding. *Pacific-Asia Conf. on Knowledge Discovery and Data Mining*, Singapore.

Tan, H., Dillon, T.S., Hadzic, F., Chang, E., 2006c. Razor: mining distance constrained embedded subtrees. *IEEE ICDM'06 Wkshp on Ontology Mining and Knowledge Discovery from Semistructured documents*, China.

Zaki, M.J., Wang, J.T.L., Toivonen, H.T.T., 2003. Data Mining in Bioinformatics: report on BIOKDD'03'. *SIGKDD Explorations*, vol. 5, no. 2, pp. 198-200.

PROCESS AND E-SERVICE CUSTOMIZATION
For Coordination in Healthcare Networks

Günter Schicker, Carolin Kaiser and Freimut Bodendorf
Department of Information Systems, University of Erlangen-Nuremberg, Lange Gasse 20, 90403 Nuremberg, Germany
guenter.schicker@wiso.uni-erlangen.de, carolin.kaiser@wiso.uni-erlangen.de, bodendorf@wiso.uni-erlangen.de

Keywords: Healthcare networks, processes, case based reasoning, coordination, customization, e-services.

Abstract: Coordination in healthcare networks becomes increasingly important. A process-oriented coordination approach is introduced which enhances integrated care scenarios by an IT-driven coordination of interorganizational treatment processes – the concept of process-based e-service logistics. The allocation of e-services is based on a model describing services and coordination tasks between roles in a healthcare network. The underlying system's architecture is presented which implements process-based e-service logistics by designing and executing individual treatment processes, identifying coordination tasks between network actors and dynamically allocating e-services. A solution for automated individualization of processes and e-services based on Case Based Reasoning (CBR) technology is discussed.

1 HEALTHCARE NETWORKING

The healthcare industry is an important economic sector in Germany causing annual expenses of about 230 billion euros and employing more than 4.2 million people. Starting point of the research project was an empirical study addressing German and Swiss ambulant healthcare networks (healthcare network managers as well as physicians) to evaluate the maturity of healthcare network organizations regarding strategy, processes and information technology (Schicker et al., 2006a). The empirical study reveals that 81 percent of the respondents expect that networking in the healthcare industry will increase in the next three to five years. Moreover, 88 percent of the survey participants agree that the demand for coordination and IT-support in healthcare networks is going to rise in the future (Schicker et al., 2006a: 21). An important driving force of networking is the concept of integrated care which is often associated with the following instruments:

- Intersectorial cooperation (Ramming, 2004: 147; Mühlbacher, 2002: 65)
- Financial responsibility (e.g. capitation)
- Coordination and control of medical treatment processes (Mühlbacher, 2002: 66)
- Information integration (Ramming, 2004: 147).

Whereas many research projects deal with the integration of health data this project concentrates on the coordination and control of interorganizational processes within healthcare networks by providing patients and suppliers with a customized set of electronic services. To analyze the requirements the research team cooperates with the healthcare network "Qualitäts- und Effizienzgemeinschaft Nürnberg-Nord" which is organized as a gatekeeper system. The integrated care contract spans ambulant, clinical and home care service providers and is financed by a full capitation model (Wambach et al., 2005: 13).

2 PROCESS-ORIENTED COORDINATION

Process-oriented coordination is seen as one important way to enable integrated care scenarios, to enhance patient satisfaction and to reduce costs of treatment processes (Güntert, 2004: 100ff). Schmalenbach argues for process-oriented control especially to cope with the problem of managing multiple interfaces (Schmalenbach, 1908/09; 211f.). Based on that belief the research project focuses on the treatment process from a cooperative view regarding the patient's way throughout the whole healthcare network.

2.1 Process Characteristics and Requirements

Process-oriented coordination faces important challenges to cope with. Therefore the characteris-

tics of healthcare network processes and their requirements will be discussed next. To transfer the general tasks and principles of coordination to the healthcare domain, it has been analyzed which processes (e.g. management, medical treatment and support processes) and coordination tasks exist within healthcare networks (Schicker, 2006b: 39).

2.1.1 Process Characteristics

Table 1 depicts characteristics of interorganizational treatment processes that need to be considered when requirements for the IT-support in healthcare networks are defined (Schicker et al., 2006b: 39).
Especially the uniqueness of one patient´s treatment process (Müller-Mundt, 2001: 95) and the high degree of volatility are key characteristics which have to be considered when supporting healthcare network processes by information technology. During their execution processes have to be modified, detailed and customized to the patient´s needs depending on his individual treatment context (cybernetic model of medical treatment) (Prokosch, 2007).

Table 1: Characteristics of treatment processes.

• unique • requiring intensive coordination and information • cross-sectorial • important and risky	• stepwise changing and volatile, requiring in-depth knowledge • involving numerous participants • long-lasting and complex

2.1.2 Requirements

The individual characteristics of each patient, the high degree of volatility during the real-time execution of the process instance are crucial challenges when supporting modelling, adaptation and execution of individual processes by information technology (Purucker et al., 2007).

For that reason process models should be adapted to the individual needs of the patients (Haas, 2005: 553ff). To avoid modelling efforts a new automated process design is needed.

Moreover, members of the healthcare network (especially gatekeepers) should be able to modify the individual process model easily as soon as new information about the state of treatment or illness exists (Schwarz et al., 2001: 10; Remus, 2002: 115). Finally, e-services (e.g. information and application services) must be linked to individual process activities in a flexible way.

3 PROCESS-BASED E-SERVICE LOGISTICS

3.1 Basic Principles

The concept of process-based e-service logistics (PEL) aims to support the coordination of healthcare network processes by providing patients and healthcare suppliers with a customized set of electronic services. E-services are software components which encapsulate functions (e.g. logic or data centric services) in a coarse-grained manner, e.g. using web services as technical representation (Krafzig et al., 2004: 70ff). The e-service requirements regarding information and coordination in networks are derived from customized process models. They result in a process-based e-service logistics model executed by a process management platform supporting the coordination of individual treatment processes by providing network participants with e-services.

3.2 Architecture Overview

The proof of the concept described above is being implemented and called Individual Value Web System (IVWS). IVWS is a process platform supporting configuration, execution and control of processes across healthcare networks. The system architecture follows a four-tiered approach (see figure 1).

3.2.1 Presentation Layer

Application front ends initiate and control all activities of the IVWS. Typical application front ends are graphical user interfaces enabling direct interaction of users with the system (WebParts implemented with C# using the Sharepoint Portal Server). Within the research project role-specific process portals for patients, service providers and network managers are implemented (Schicker et al., 2005: 7ff).

3.2.2 Customization and Flow Control Layer

This layer consists of three components including features for orchestration and execution of web services.

- Service bus: This component ensures the execution of web service orchestrations and provides functionalities of a service bus: connectivity, integration and communication services, process orchestration and execution (Krafzig et al., 2004: 65).
- Process and e-service customization: An important issue of PEL is how to adapt and

model one patient's treatment process and the underlying e-services needed for supporting the process (for details see section 4).

- Meta orchestration server (MOS): MOS enables the configuration, execution and monitoring of treatment processes. MOS represents a process engine which acts as a broker between the front ends and the e-services invoked during process execution. It also initiates the execution of web service orchestrations by the BizTalk Server and integrates process proposals created by the customization component for process individualization (for details see Purucker, 2007).

3.2.3 Application Layer

This layer contains all e-services needed for execution and support of treatment processes. E-services of a SOA are software components encapsulating business functions in a coarse-grained manner. An e-service consists of a service contract, a service interface and an implementation. The implementation is the physical representation of the required business logic and the relevant data (programs, configuration data and database) (Krafzig et al., 2004). E-services include IVWS-specific services (e.g. the customization component described below) as well as third party web services and adapters for the integration of existing information systems.

3.2.4 Data Layer

A MS SQL-based process and e-service repository provides features for searching and applying process models and e-services, e.g., physical location, service provider, charge fee, technical constraints, security data and service level agreements.

4 PROCESS AND E-SERVICE CUSTOMIZATION

4.1 Traditional Methods

There are many approaches which enable the customization of processes in general and which can be applied for the customization of treatment processes and e-services. Following the classic approach, each treatment process model and each e-service is manually developed from scratch for each context with respect to guidelines (Lang, 1997: 29). The steady process of creating customized treatment processes and e-services is very time-consuming and inhibits the reuse of experiences. To overcome this disadvantage new concepts have been developed. The process modelling approaches based on reference models (vom Brocke, 2003: 31ff), process skeletons (Remme, 1997:114ff) and process modules (Lang 1997: 4ff) enable the reuse of treatment process knowledge. The approaches for creating e-services by using design patterns, frameworks and libraries allow the reuse of e-service knowledge. However, these concepts require a time-consuming search and adaptation to the current context.

The disadvantage of existing approaches (e.g Wargitsch et al., 1997: 3ff; Rupprecht, 2002: 67ff) is that only one of the two steps (search and adaptation) is automated whereas the other step must be executed manually. An approach is needed which enables an automatic search and adaptation (Schicker et al., 2007). Here, such a system based on Case Based Reasoning (CBR) is introduced.

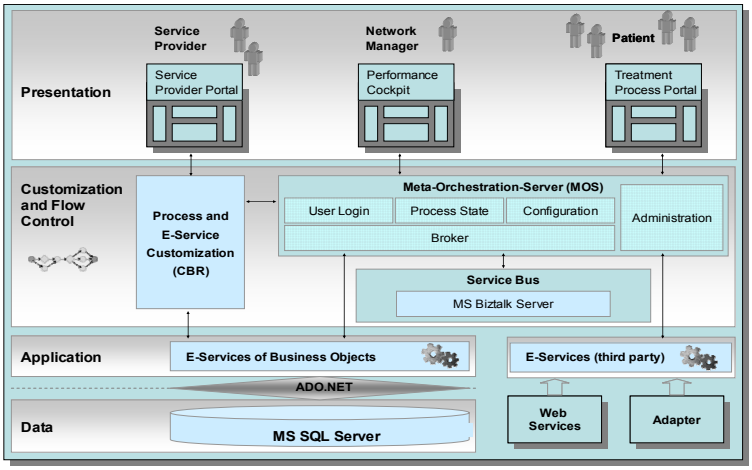


Figure 1: IVWS architecture.

4.2 Case Based Reasoning

4.2.1 Introduction

CBR is a problem solving paradigm of Artificial Intelligence. It solves new problems based on past experiences saved in form of cases in the case base. Each case (patient record) consists of a problem description (patient context) and a solution (treatment process and its e-services). In order to propose a treatment process and e-services for a given patient, CBR searches its case base for the patient record which is most similar to the given patient record and adapts its treatment process with e-services to fit the given case.

Besides fulfilling the requirements of automated search and adaptation, CBR offers several advantages (Nilsson and Sollenborn 2004: 178). First of all, it resembles the physicians' cognitive process of recalling former patients and reusing past experiences. This resemblance does not only prove the force of the approach but also leads to high user acceptance. Furthermore, the collection of patient records can easily be integrated in a CBR system as a case base. Moreover, the reuse of patient records provides an efficient reasoning mechanism which does not require solving each problem from scratch.

The functionality of a CBR system can be divided into four main phases which form the CBR cycle: retrieve, reuse, revise, retain phase (Aamodt and Plaza 1994: 46). The CBR cycle for customizing treatment processes and e-services is illustrated in figure 2 and described in the following sections.

4.2.2 Representation of Patient Records

Each patient record (case) contains the patient context (problem) as well as its treatment process and e-services (solution). While the patient context specifies the patient and his disease pattern, the treatment process and its supporting e-services characterize the therapy of the patient.

Attribute-value-vectors are used to represent the patient context. According to this type of representation, the patient context is specified by a vector of attribute values. In order to reflect the characteristics of different diseases, the set of attributes has to be defined for each disease separately.

Treatment processes and e-services can consist of several elements which are represented in an object-oriented manner. They are specified by a set of instances of classes (e.g. treatment process, service, coordination task, patient task and e-service).

4.2.3 Retrieve Phase

In the retrieve phase a health care provider passes the attribute values of the patient context to the system and requests a treatment process. In order to fulfil this request, the system searches for the patient record on the case base whose patient context is most similar to the context of the given patient. The search for a similar case requires the definition of similarity measures and search algorithms.

The similarity calculation of two patient contexts is based on local and global similarity measures (Stahl, 2003: 50ff). Local similarity measures determine the similarity between one attribute-value of the query patient record and the patient record in the case base, whereas global similarity measures calculate the aggregated similarity of all attribute-values. Different types of local similarity measures are used depending on the type of the attribute.

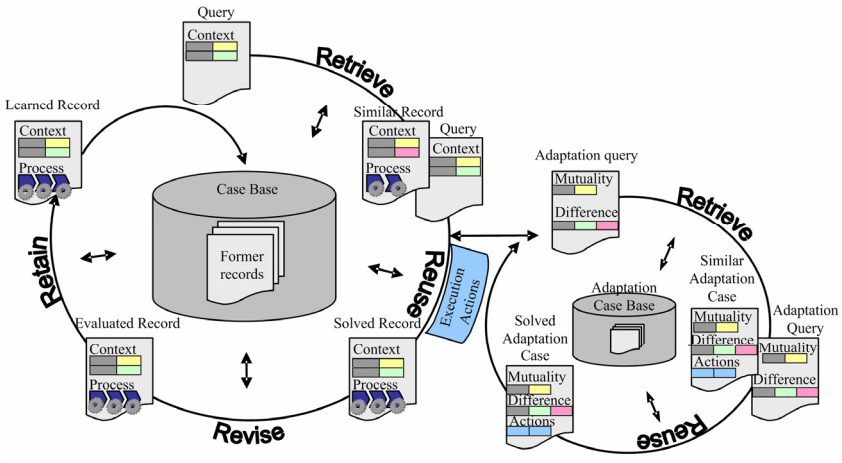


Figure 2: CBR system for customizing treatment processes and e-services.

For determining the similarity of nominal attributes, similarity tables are appropriate. They assign a similarity value to each combination of attribute values of the query patient context and the patient record in the case base. In order to calculate the similarity of metric attributes threshold-based, linear, exponential and sigmoide functions can be applied.

A weighted sum of the local similarities is chosen as global similarity measure. The weights reflect the relevance of the attributes. In order to reduce the high effort for the manual definition of similarity measurements, a learning algorithm for simplifying and optimizing the global similarity measurement has been implemented.

For searching the most similar patient records, sequential search and knowledge-poor indexing based on an extended k-d-tree (Wess, 1995: 163ff, 209ff) are provided (for details see Kaiser, 2008).

4.2.4 Reuse Phase

The reuse phase receives as input the record of the most similar patient which was found during the retrieve phase and aims at adapting it to fit the new patient record. The adaptation for the customizing of treatment processes consists of four successive steps: copy adaptation or compositional adaptation, substitutional adaptation, structural adaptation and consistency assurance.

The first adaptation step takes different turns depending on the phase of the treatment process. If a new treatment process should be created the treatment process of the similar patient record can be copied. However, if an existing treatment process must be extended, elements of the similar patient record are added. The substitutional adaptation replaces the attribute values of treatment elements and e-service elements adopted from the similar patient record according to rules. The structural adaptation modifies the structure of the requested treatment process and its e-services by deleting, adding and rearranging elements. It is realized by an additional adaptation-based CBR system (see figure 2). The adaptation cases of the adaptation-based CBR system describe which modification actions (solution) should be taken when certain differences and mutualities of the patient context (problem) between query and record in the case base occur (for details see Kaiser, 2008). The consistency assurance aims at improving the consistency of the created treatment process and its e-services.

4.2.5 Revise Phase

During the revise phase, the execution of some treatment steps and e-services takes place. In order to support the execution, the treatment process and its e-services are transformed into an XML representation and passed on to the meta-orchestration server for further processing.

4.2.6 Retain Phase

The retain phase is the last phase of the CBR cycle and belongs to the case base maintenance. The aim of the case base maintenance is to detect environmental changes which could decrease the quality and efficiency of the case base and execute counteractive measures. Besides the retain phase, the restore phase and the review phase are also part of the case base maintenance (Roth-Berghofer 2002: 55ff). The retain step uses intra-case quality measures to decide on the insertion of a patient record and is called each time a patient record has passed the revise phase. The review phase and restore phase are executed periodically. Hereby all patient records in the case base are checked with the aid of inter-case quality measures and redundancy measures and added or deleted accordingly.

5 CONCLUSIONS

The importance of coordination and IT-support in healthcare networks is increasing steadily. In order to enhance the quality and efficiency of interorganizational treatment processes, an IT-supported process-oriented coordination approach for healthcare networks has been realized. The approach supports the configuration, execution and control of processes and supporting e-services in healthcare networks.

This paper focuses on the configuration of treatment processes and e-services. Uniqueness and volatility are key characteristics of treatment processes. Treatment processes and e-services have to be configured whenever new information on the status of the patient becomes available. In order to reduce the manual effort for searching and adapting treatment processes and e-services an automated system based on CBR has been developed.

The definition and update of the patient context attributes, treatment process steps and e-services is time-consuming. However, it enables to formalize and expatiate on medical treatment knowledge. The approach offers basic functionality and can be extended by further methods such as the prediction of unknown attribute values of the patient context.

REFERENCES

Aamodt, A., Plaza, E., 1994. Case Based Reasoning: Foundational Issues, Methodological Variations and System Approaches. *AI Communications* 7 (1), 39-59.

Güntert, B. J., 2004. Integration und Kooperation im Gesundheitswesen – ein Plädoyer für patientenorientierte Strukturen und Verhaltensweisen. In: *Busse R. et al. (Ed.): Gesundheitswesen – vom Heute ins Morgen*. WuV Universitätsverlag, Wien, 100-109.

Haas, P., 2005. *Medizinische Informationssysteme und Elektronische Krankenakten*. Springer, Berlin.

Kaiser, C., 2008. Case based Reasoning for Customizing Treatment Processes. To appear in Lazakidou, A. A., Siassiakos, K. M.. *Handbook of Research on Distributed Medical Informatics and E-Health*. Idea.

Mühlbacher, A., 2002. *Integrierte Versorgung: Management und Organisation*. Hans Huber, Bern.

Müller-Mundt, G., 2001. Patientenedukation zur Unterstützung des Selbstmanagements. In: *Hurrelmann, K., Leppin, A. (Ed.): Moderne Gesundheitskommunikation: vom Aufklärungsgespräch zu E-Health*. Hans Huber, Bern, 94-106.

Krafzig, D., Banke, K., Slama, D., 2004. *Enterprise SOA, Service-Oriented Architecture Best Practices*. Prentice Hall PTR, Maryland.

Lang, K., 1997: *Gestaltung von Geschäftsprozessen mit Referenzbausteinen*. Deutscher Universitätsverlag, Wiesbaden.

Prokosch, H. U., 2007. Kybernetisches Modell ärztlichen Handelns. In: *Vorlesungsunterlagen zur Einführung in die Medizinische Informatik für Informatik-Nebenfachstudierende.* Retrieved February 18, 2007, from http://www.imi.med.uni-erlangen.de/lehre/ws 0506/medinfgrund_06.pdf.

Purucker, J., Schicker, G., Affes, F., Bodendorf, F., 2007. Meta-Orchestration of E-Services for Process-oriented Coordination in Healthcare Networks. In: *Lecture Notes in Informatics (LNI), Gesellschaft für Informatik (GI)*. Köllen Verlag, Bonn.

Ramming, J., 2004. Integrierte Gesundheitsversorgung. In: *Jähn, K., Nagel, E. (Ed.): e-Health.* Springer, Heidelberg, 147-151.

Remme, M., 1997. *Konstruktion von Geschäftsprozessen – Ein modellgestützter Ansatz durch Montage genetischer Prozesspartikel*. Gabler Verlag, Wiesbaden.

Remus, U., 2002. Integrierte Prozess- und Kommunikationsmodellierung zur Verbesserung von wissensintensiven Geschäftsprozessen. In: *Abecker, A et al. (Ed.): Geschäftsprozessorientiertes Wissensmanagement.* Springer, Berlin, 91-122.

Roth-Berghofer, T., 2002. *Knowledge Maintenance of Case Based Reasoning Systems - The SIAM Methodology*. University of Kaiserslautern, Dissertation.

Rupprecht, C., 2002. *Ein Konzept zur projektspezifischen Individualisierung von Prozessmodellen*. Universität Karlsruhe, Dissertation.

Schicker, G., Bodendorf, F., 2005. Portalunterstützte Behandlungspfade in Gesundheitsnetzen: Prozesse - Architektur - dynamische Navigation. In: *Cremers, A. B. et al. (Ed.): Lecture Notes in Informatics (LNI), Gesellschaft für Informatik (GI)*. Köllen, Bonn, 7-11.

Schicker, G., Bodendorf, F., 2006. Process-based E-Service-Logistics for Healthcare Networks. In: *Stormer H., Meier A., Schumacher M. (Ed.): Lecture Notes in Informatics (LNI) - Proceedings. Gesellschaft für Informatik (GI)*. Köllen, Bonn, 37-46.

Schicker, G., Kaiser, C., Bodendorf, F., 2007. Individualisierung von Prozessen und E-Services mithilfe von Case Based Reasoning. In: Oberweis, A. et al. (Ed.): *eOrganisation: Service-, Prozess-, Market-Engineering/Band 1*. Universitätsverlag Karlsruhe, Karlsruhe, 713-730.

Schicker, G., Kohlbauer, O., Bodendorf, F., 2006. *Praxisnetz-Studie 2006 - Status Quo, Trends & Herausforderungen*. Arbeitspapier Wirtschaftsinformatik II Nr. 01/2006, Universität Erlangen-Nürnberg. Nürnberg.

Schmalenbach, E., 1908/09: Über Verrechnungspreise. In: *Zeitschrift für handelswissenschaftliche Forschung 3*, 165-185.

Schwarz, S., Abecker, A., Maus H., Sintek, M., 2001. *Anforderungen an die Workflow-Unterstützung für wissensintensive Geschäftsprozesse*. Retrieved October 29, 2006, from http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-37/Schwarz.pdf.

Stahl, A., 2003. *Learning of Knowledge-Intensive Similarity-Measures in Case Based Reasoning*. Universität Kaiserslautern, Dissertation.

Nilsson, M., Sollenborn, M., 2004. Advancements and Trends in Medical Case-Based Reasoning: An Overview of Systems and System Development. In *Proceedings of the FLAIRS Conference*, AAAI Press, 178ff.

Vom Brocke, J., 2003. Referenzmodellierung – Gestaltung und Verteilung von Konstruktionsprozessen. In Becker et al. (Eds.). *Advances in Information Systems and Management Science*. Logos, Berlin.

Wambach, V., Lindenthal, J., Frommelt, M., 2005. *Integrierte Versorgung - Zukunftssicherung für niedergelassene Ärzte*. Ecomed, Landsberg, 2005.

Wargitsch, C., Wewers, T., Theisinger, F., 1997: WorkBrain: Merging Organizational Memory and Workflow Management Systems. In *Workshop ‚Knowledge Based Systems for Knowledge Management in Enterprises' im Rahmen der 21. KI-Jahrestagung*. Freiburg.

Wess, S., 1995: *Fallbasiertes Problemlösen in wissensbasierten Systemen zur Entscheidungsunterstützung und Diagnostik*. Infix Verlag, Sankt Augustin.

A SMART MEDICINE MANAGER DELIVERING HEALTH CARE TO THE NETWORKED HOME AND BEYOND

An Overview of the iCabiNET System

Martín López-Nores, José J. Pazos-Arias, Jorge García-Duque and Yolanda Blanco-Fernández
Department of Telematics Engineering, University of Vigo, Spain
{mlnores, jose, jgd, yolanda}@det.uvigo.es

Keywords: Remote health care, smart medicine packaging, residential networks, semantic reasoning.

Abstract: Misuse of prescription and over-the-counter drugs is a growing problem that impinges heavily on the well-being of people and the economics of public health systems. Most commonly, misuses arise from forgetfulness or lack of information about drugs and their interactions, hence there is much place for solutions to automatically monitor medicine intake, issue reminders and deliver medical advice. This paper presents a system that accomplishes these tasks by harnessing recent advances in smart medicine packaging, residential networks and semantic reasoning. Such a combination yields a medicine manager featuring great precision in drug monitoring, plus unprecedented capabilities to reach the users and provide them with valuable information.

1 INTRODUCTION

Recent statistics reveal that the misuse of prescription and over-the-counter drugs is becoming a major problem as life expectancy increases and the range of medications grows, to the point of being as dangerous and costly as many illnesses (Sullivan et al., 1990; Downey et al., 2000). To ground this significance in numbers, consider the following USA facts, retrieved from (American Heart Association, 2007; Akram, 2000; Office of Applied Studies, 2005):

- 50% of filled prescriptions are taken incorrectly. In the worst extreme, 65% of the elderly fail to comply with their medication regimens, with 26% of those errors being potentially serious.
- 23% of nursing home admissions are due to abuse or non-compliance, costing $31.3 billion per year and affecting 380,000 people. The same happens with 10% of hospital admissions, costing $15.2 billion and affecting 3.5 million people.
- $75 billion are annually spent on preventable hospitalizations due to medication misuse, plus other $30 billion on additional medications prescribed after non-compliance.
- 125,000 deaths occur annually due to drug interactions.

As noted in (Hughes et al., 2001; MedPrompt, 2007), misuses typically arise from forgetfulness or lack of information about the different drugs available. This has raised enormous interest in developing solutions to automatically monitor medicine intake, issue reminders and deliver medical advice (Bricon-Souf and Newman, 2007). Some precedents for this idea can be found in (Wan, 1999; Ho et al., 2005), with embedded systems that employed RFID[1] devices to recognize medicaments and weighing scales to guess the doses available. However, those systems are impractical due to the following drawbacks:

- The weighing scales require the users to pick one medicament at a time and put it back before picking another. This is a cumbersome discipline to follow in many cases, entailing a clear risk of monitoring imprecision.
- The means available to reach the users are very limited, merely consisting of embedded screens, lights or alarms. Thus, it is not possible to prevent forgetfulness if the user happens to be out of home, or simply in a room where he/she cannot see the lights or hear the alarms.
- Finally, the previous systems rely on the assumption that people use a medicine cabinet as the

[1] RFID: Radio Frequency IDentification.

only place for medication keeping. Nevertheless, polling data reveals that this may not be true in as many as 90% of the cases, as people tend to store their medicines in various places around the house (Fishkin and Consolvo, 2003).

With these problems in mind, we introduce in this paper a system, called the iCabiNET, that tackles the aforementioned issues by integrating recent advances in various areas of research. The basic ideas behind this system are explained in the overview of Section 2. After that, Section 3 describes two usage scenarios to illustrate the benefits and the potential uses of our approach. Section 4 provides technical details of an implementation capable of realizing those scenarios. Finally, Section 5 includes a summary of conclusions and motivates future work.

2 SYSTEM OVERVIEW

As shown in Fig. 1, we have conceived the iCabiNET as a new element of a residential network, ready to communicate directly with other appliances installed in a house, and with the outside world through a residential gateway. Within this setting, the operation of the system consists of two major steps, to be detailed in the following subsections:

- Gathering information about available drugs and doses.
- Processing that information to identify and react to actual or potential misuses.

2.1 Gathering Information

In what concerns the gathering of information about available medicines and doses, the iCabiNET can coexist with any of the previous monitoring solutions (Wan, 1999; Floerkemeier and Siegemund, 2003; Ho et al., 2005), given that they propagate data over the residential network. Furthermore, we have introduced support for the *smart packaging technologies* currently promoted by stakeholders of the pharmaceutical industry. As explained in (Goodrich, 2006; Harrop, 2006), the idea is to integrate RFID devices and different types of sensors with the packaging of the medicines, to allow tracking not only medicine names, but also the doses available with no additional equipment. A common example is that of *smart blister packs*, which record the removal of a tablet simply by breaking an electric flow into the RFID's integrated circuit; other possibilities exist for liquid medicines, ointments and so on.

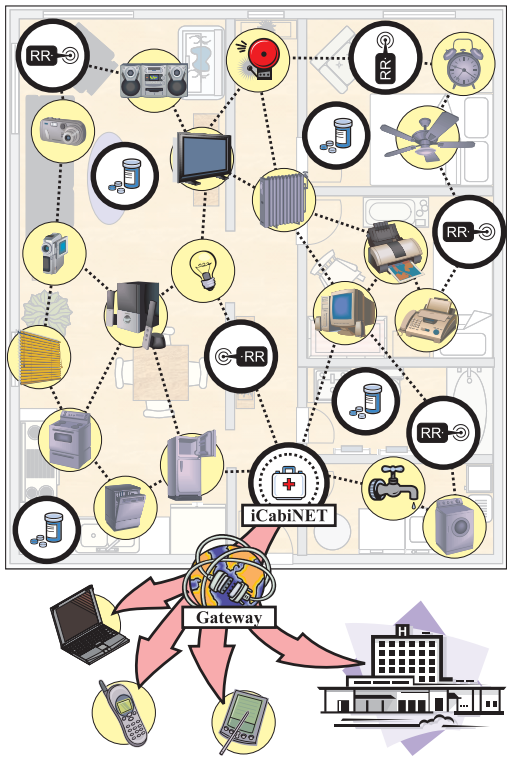


Figure 1: The iCabiNET system in a residential network.

Smart packaging enables the iCabinet to monitor the intake of drugs from anywhere, with the highest precision, with no need to keep all drugs in a unique place, and with no risk of mistaking one drug for another. The system simply gathers information by polling RFID readers deployed around the house (the RR devices in Fig. 1) or connected to the residential network from other places (e.g. from the office).

2.2 Reacting to Misuses

Primarily, the iCabiNET is intended to enforce some *medication guidelines*, such as "*the user should take one of these tablets every 4 to 6 hours*". Accordingly, in the operational scheme of the system (see Fig. 2), there is a 'Watchdog' module devoted to continually supervising the information gathered about available medicines and doses, to check that the former remain in good condition and the latter decrease correctly with time. This module detects odd circumstances driven by *rules* that may involve user conditions (like age, gender or previous diseases), and notifies those circumstances by triggering different types of *events*.

The events are the input for a second module, the 'Actuator', to decide what *actions* will be performed to issue warnings or deliver health care information to the user. This module firstly considers generic state-

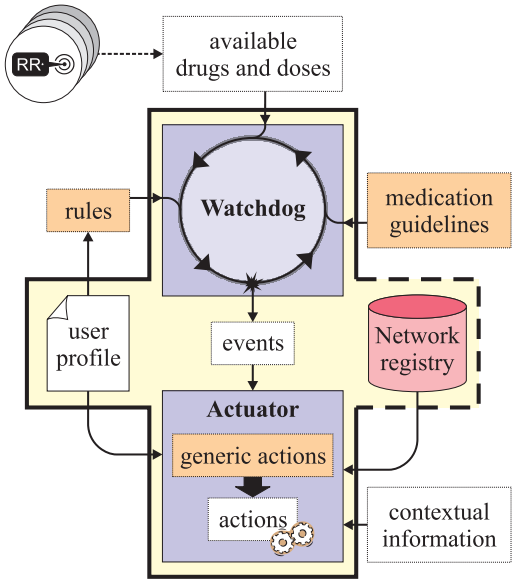


Figure 2: The operational scheme of the iCabiNET.

ments with no liaison to specific appliances, such as those of Table 1. Then, it instantiates those statements on demand, using the appliances it finds most convenient. In doing so, the 'Actuator' takes into account data from the user's profile, contextual information provided by external devices (e.g. about whether the user is sleeping, watching TV or out of home) and descriptions stored in a *network registry* of the appliances connected to the residential network and the operations they can do.[2] Thus, for example, a "*warn the user*" action can be automatically made to trigger an alarm clock, to interrupt a TV program and display some message on screen, or to make a telephone call. The enormous range of possibilities enabled by the residential network to reach the user is precisely the point that makes the iCabiNET most advantageous with regard to previous systems.

Orchestrating appliances as explained above requires the iCabiNET to take a great number of decisions that cannot be determined beforehand. Noticeably, this *intelligence* is not catered for by the current residential network standards (Baxter, 2005), because they merely provide for executing pre-compiled programs (usually referred to as *bundles*). Thus, it would be necessary to write different versions of the same behavior for all the possible configurations of devices in and out of home, even considering the different ways to invoke the same operations for appliances assembled by different manufacturers. Such an approach would obviously exhibit limited flexibility

[2]The registry can reside in the iCabiNET or in any other device permanently connected to the residential network, most typically the residential gateway.

Table 1: Some events types and actions they might trigger.

Event	Action
Oblivion	"*Wait up to 90 minutes before reminding the user*"
Expiration	"*Deliver increasingly serious warnings day after day*"
Depletion	"*Arrange an appointment with the doctor to get a new prescription*"
Interaction	"*Recommend an innocuous combination of drugs with the same effects*"
Discontinuation	"*Restart the medication at a lower dose*"
Abuse	"*Warn the local authorities*"

and severe scalability problems. The iCabiNET's solution to these questions builds upon two main ideas:

- The first idea is to deliver *virtual bundles* containing no implementation, but rather process flows that arrange medication guidelines, user conditions, rules, events and generic actions in semi-formal constructs.
- The second idea is to enhance the network registry with mechanisms from the Semantic Web (Antoniou and van Harmelen, 2004), using *ontologies* as unique conceptualizations of what the different appliances can do and how: operations, input and output parameters, quality attributes, etc.

As shown in Fig. 3, when a virtual bundle appears in the system, the iCabiNET creates one *implementation bundle* to supervise the occurrence of events in the corresponding process flow. Then, when it is time to perform some actions, the implementation bundle uses matching techniques like those of (Paolucci et al., 2002; Fujii and Suda, 2005) to find the most suitable appliances at the moment (A, B, D and E in Fig. 3) and start invoking their operations in the specified order. This approach promotes openness and interoperability, making it possible to deliver the same virtual bundles to everybody, regardless of the particular appliances owned by each user. Furthermore, nothing has to be re-programmed when any element changes, and it is even possible to incorporate newly-invented devices and functionalities with a simple update of the ontologies.

Following the commented scheme, several virtual bundles are preloaded in the iCabiNET to deal with common tasks, such as checking interactions between available drugs by accessing remote data bases, or downloading process flows to drive the monitoring of new drugs acquired by the user. Other virtual bundles

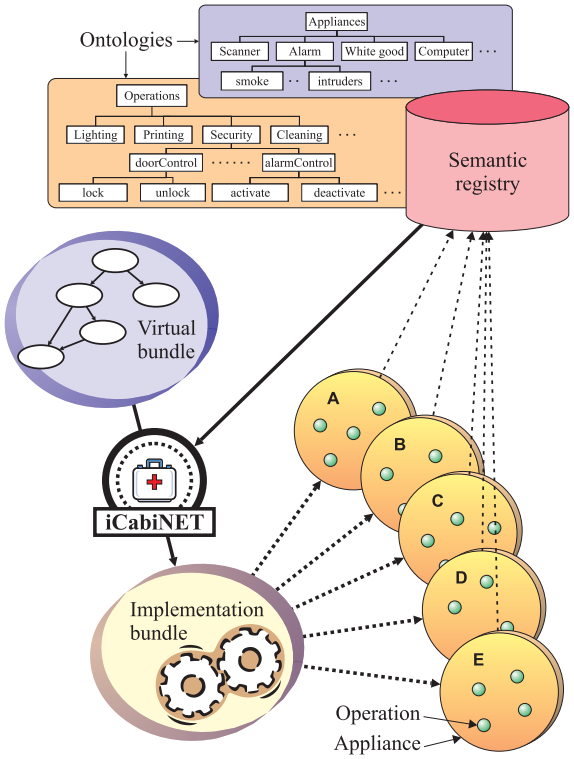


Figure 3: Activating appliances from virtual bundles.

can be entered by the user or by authorized external entities, like health institutions (they know what the user needs) or pharmaceutical companies (they know the best way to take their products).

3 USAGE SCENARIOS

Having explained the essentials of the iCabiNET, we now describe two usage scenarios to illustrate the range of functionalities it can deliver. The technologies employed to make these scenarios possible are described in Section 4.

3.1 Scenario #1

Ann is having breakfast before going to work, and switches on the radio to hear the first news of the day. When she is about to turn off the apparatus, the iCabiNET reminds her of the medicines she should carry, playing a pre-recorded message. Later, following the prescription issued by the doctor (loaded into the iCabiNET directly from the health center), Ann receives an SMS message in her mobile phone every three hours to remember taking her drugs. At the end of the day, when Ann is back in her house, the iCabiNET attempts to check that the available doses have decreased as expected, but it turns out that Ann has left them behind. In this case, the iCabiNET rings the in-home telephone to ask Ann whether she has taken the medicines correctly; she replies affirmatively by pressing the asterisk key. During the night, the alarm clock in Ann's bedroom will be responsible for waking her up when it is time for new doses.

3.2 Scenario #2

While having a walk outside, Bob decides to buy an over-the-counter drug to treat his allergy to pollen. Afterwards, when he enters his house, the iCabiNET records the tablets he has bought, and automatically downloads medication guidelines for adults from the manufacturer. A few days later, Bob is watching TV in the living room. When it is time to take a pill, the iCabiNET pops up a reminder on the screen, indicating the drug's commercial name, a photograph of its packaging and the recommended dose. For the best comfort, the system starts flashing the lights of the room where Bob had left the tablets the last time he took one. As Bob takes a new dose, the iCabiNET finds that the pills are running out; so, when he sits back in the sofa, he is faced with an interactive TV application that he can use to buy new supplies from an online drugstore. Bob uses the remote control to enter shipping and payment details, and takes the opportunity to buy some throat lozenges he likes.

4 PROTOTYPE IMPLEMENTATION

In order to assess the feasibility of the approach described in Section 2, we have developed and tested a prototype of the iCabiNET taking the scenarios of Section 3 as a reference of the functionalities it should provide. It is worth noting that the system needs not be a standalone device, hence we built it as a software package to run on any device that is permanently connected to a residential network. Within this perspective, we strove to employ standard technologies and open-source software packages.

As the basis for our implementation, we chose the framework proposed by OSGi (OSGi Alliance, 2005), which is nowadays the most popular standard for residential networks. This platform is advantageous for various reasons: (i) it supports different widespread protocols for secure and non-secure communication among appliances in home and outside, (ii) it defines a cooperative model where appliances can dynamically discover and invoke the operations provided by others, and (iii) it enables remote management of the

appliances and the operations they provide. From among other possibilities, we implemented the iCabiNET as an OSGi bundle using the open software packages from the OSCAR[3] project, because they are particularly well documented.

Just like the other residential network standards, OSGi provides no support for orchestrating appliances according to virtual bundles as described in Section 2.2. To this aim, inspired by the work of (Slomiski, 2006), we opted to borrow solutions from the most mature related field of research: Web Services. Therefore, we express the process flows using the BPEL language (Juric, 2006), which provides constructs to describe arbitrarily complex processes, focusing on the invocation of operations and the flow of control between them. There are many tools supporting this language, offering plenty of facilities to create, edit and execute process flows. We have developed the core of the iCabiNET over the ActiveBPEL[4] engine, introducing the following enhancements:

- The BPEL flows can include declarative rules written in Jess[5] to drive the generation of events related to the intake of medicines. Accordingly, the 'Watchdog' module of Fig. 2 incorporates a Jess execution environment, which is the only protected part of our implementation —it is only free for academic purposes.
- The actions in the BPEL flows include concepts of the SOUPA and GUMO ontologies presented in (Chen et al., 2004; Heckmann et al., 2005), which are also used in a semantic registry maintained by the iCabiNET itself. With those bases, plus the context-aware features of (Gu et al., 2004; Zhang et al., 2005), the 'Actuator' module of Fig. 2 uses the Protégé OWL API[6] to apply the same semantic matching mechanisms we designed for (Díaz-Redondo et al., ress) —the reasoning abilities are not linked to any specific domain of application.

As regards the interactive applications presented to the user to realize actions in the virtual bundles (e.g. the online drugstore of Scenario #2), the authors of (Ramos-Cabrer et al., 2006) proposed a way to merge OSGi with the MHP standard of applications for Digital TV (Digital Video Broadcasting, 2003). We have extended those mechanisms to support applications for PCs and mobile devices as well, and to handle descriptions of those applications also in terms of the SOUPA and GUMO ontologies.

All the code aforementioned pieces of the iCabiNET are written in Java, and so we used this language for the code to glue those elements together. That code was executed by a J2ME virtual machine running on a residential gateway. Out of the iCabiNET itself, in what concerns the gathering of information about the availability and intake of medicines, we completed the settings for our trials with purpose-built smart blister packs, because smart packaging technology is not yet available in retail drugs (only in clinical trials). Besides, we developed our own RFID readers for those blister packs, since we did not find suitable alternatives in the market ready to work within an OSGi network.

5 CONCLUSIONS

Technology may be an important aid to fight the worrisome health problems and the increasing economic costs due to the misuse of prescription and over-the-counter drugs. With this vision in mind, we have designed and implemented the iCabiNET system, which is the first outcome of a multidisciplinary research effort to put technology to the service of better medicament monitoring and management.

The iCabiNET can be regarded as a powerful extension of previously-existing solutions to monitor medicine intake (e.g. the embedded systems of (Wan, 1999; Floerkemeier and Siegemund, 2003; Ho et al., 2005)), introducing the ability to issue warnings and deliver medical advice using any appliance connected to a residential network. Eventually, when the smart packaging technologies finally take up the market, the iCabiNET approach will actually replace that of previous systems, providing the greatest flexibility and precision in monitoring, limitless possibilities to reach the user using whichever technological means, and unprecedented capabilities to interact with health institutions, drug manufacturers and retailers.

Also, the iCabiNET can be seen as introducing support to monitor the intake of drugs in previously-existing platforms to provide health care information through Internet-enabled personal computers (Roine et al., 2001), Digital TV (Simonov et al., 2007) or mobile devices (Komninos and Stamou, 2006). In this regard, we can emphasize the integrated and technology-neutral solution achieved with the semantic reasoning features, which render an open environment with feasible and scalable exploitation models.

[3] http://oscar.objectweb.org
[4] http://www.activebpel.org
[5] http://herzberg.ca.sandia.gov/jess
[6] http://protege.stanford.edu

ACKNOWLEDGEMENTS

This work has been supported by the Spanish Ministry of Education and Science (Project TSI2007-61599).

REFERENCES

Akram, G. (2000). Over-the-counter medicines: An emerging and neglected drug abuse. *Journal of Substance Use*, 5(2):136–142.

American Heart Association (2007). http://www.americanheart.org.

Antoniou, G. and van Harmelen, F. (2004). *A Semantic Web primer*. The MIT Press.

Baxter, L. (2005). *Residential networks*. Thomson Delmar Learning.

Bricon-Souf, N. and Newman, C. (2007). Context awareness in health care: A review. *International Journal of Medical Informatics*, 76(1):2–12.

Chen, H., Perich, F., Finin, T., and Joshi, A. (2004). SOUPA: Standard Ontology for Ubiquitous and Pervasive Applications. In *MOBIQUITOUS'04, 1st Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services*, Cambridge, USA.

Díaz-Redondo, R. P., Fernández-Vilas, A., Ramos-Cabrer, M., Pazos-Arias, J. J., García-Duque, J., and Gil-Solla, A. (In press). Enhancing residential gateways: A semantic OSGi platform. *IEEE Intelligent Systems*.

Digital Video Broadcasting (2003). The Multimedia Home Platform specification. ETSI standard TS 102 812. http://www.mhp.org.

Downey, G., Hind, C., and Kettle, J. (2000). The abuse and misuse of prescribed and over-the-counter medicines. *Hospital Pharmacist*, 7(9):242–250.

Fishkin, K. and Consolvo, S. (2003). Results of medication questionnaire #1. Technical Report IRS-TR-03-002, Intel Research.

Floerkemeier, C. and Siegemund, F. (2003). Improving the effectiveness of medical treatment with pervasive computing technologies. In *Ubicomp'03 Workshops*, Seattle, USA.

Fujii, K. and Suda, T. (2005). Semantics-based dynamic service composition. *IEEE Journal on Selected Areas in Communications*, 23(12):2361–2372.

Goodrich, N. (2006). Smart packaging. http://www.pac.ca//Chapters/ONSmartPkgDownloads.html.

Gu, T., Pung, H., and Zhang, D. (2004). Toward an OSGi-based infrastructure for context-aware applications. *IEEE Pervasive Computing*, 3(4):66–74.

Harrop, P. (2006). *Opportunities in pharmaceutical RFID and smart packaging*. IDTechEx.

Heckmann, D., Schwartz, T., Brandherm, B., Schmitz, M., and von Wilamowitz-Moellendorff, M. (2005). GUMO: The General User Model Ontology. In *UM'05, 10th International Conference on User Modeling*, Edinburgh, UK.

Ho, L., Moh, M., Walker, Z., Hamada, T., and Su, C.-F. (2005). A prototype on RFID and sensor networks for elder healthcare. In *SIGCOMM'05 Workshops*, Philadelphia, USA.

Hughes, C. M., McElnay, J. C., and Fleming, G. F. (2001). Benefits and risks of self medication. *Drug Safety*, 24(14):1027–1037.

Juric, M. (2006). *Business Process Execution Language for Web Services*. Packt Publishing.

Komninos, A. and Stamou, S. (2006). HealthPal: An intelligent personal medical assistant for supporting the self-monitoring of healthcare in the ageing society. In *UbiHealth'06, 4th International Workshop on Ubiquitous Computing for Pervasive Healthcare Applications*, Orange County, USA.

MedPrompt (2007). http://www.medprompt.com.

Office of Applied Studies (2005). National survey on drug use & health. http://www.drugabusestatistics.samhsa.gov.

OSGi Alliance (2005). Open Service Gateway initiative. http://www.osgi.org.

Paolucci, M., Kawamura, T., Payne, T., and Sycara, K. (2002). Semantic matching of Web Services capabilities. In *ISWC'02, 1st International Semantic Web Conference*, Sardinia, Italy.

Ramos-Cabrer, M., Díaz-Redondo, R., Fernández-Vilas, A., Pazos-Arias, J., and García-Duque, J. (2006). Controlling the smart home from TV. *IEEE Transactions on Consumer Electronics*, 52(2):421–429.

Roine, R., Ohinmaa, A., and Hailey, D. (2001). Assessing telemedicine: A systematic review of the literature. *Canadian Medical Association Journal*, 165(6):765–771.

Simonov, M., Brizzi, P., Gavelli, M., and Sciarappa, A. (2007). TV for health: One-to-one personalized content delivery in living room. In *EuroITV'07 Workshops*, Amsterdam, The Netherlands.

Slomiski, A. (2006). On using BPEL extensibility to implement OGSI and WSRF grid workflows. *Concurrency and Computation: Practice and Experience*, 18(10).

Sullivan, S., Kreling, D., and Hazlet, T. (1990). Noncompliance with medication regimens and subsequent hospitalizations: A literature analysis and cost of hospitalization estimate. *Journal of Research in Pharmaceutical Economics*, 2:19–33.

Wan, D. (1999). Magic Medicine Cabinet: A situated portal for consumer healthcare. In *HUC'99, 1st International Symposium on Handheld and Ubiquitous Computing*, Karlsruhe, Germany.

Zhang, D., Gu, T., and Wang, X. (2005). Enabling context-aware smart home with Semantic Web technologies. *International Journal of Human-friendly Welfare Robotic Systems*, 6(4):12–20.

PROCESS MODELING OF THE HEALTH SECTOR USING BPMN: A CASE STUDY

Elvira Rolón, Félix García, Francisco Ruiz, Mario Piattini
Department of Information Technologies and Systems
Indra-UCLM Research and Development Institute - University of Castilla-La Mancha
Paseo de la Universidad 4, 13071, Ciudad Real, Spain
erolon@inf-cr.uclm.es, {Felix. Garcia, francisco.ruizg, mario.piattini}@uclm.es

Luis Calahorra, Marcial García and Rafael Martin
General Hospital of Ciudad Real
Ciudad Real, Spain
lcalahorraf@sescam.jccm.es, {marcialg, rmartin}@sescam.org

Keywords: Business Process, Health Sector, Modeling, Case Study.

Abstract: The importance of the analysis, modeling and management of business process is not restricted to a specific enterprise sector. In the field of sanitary management, due to the nature of the service offered, sanitary institutions' processes are also the basis for decision making focused on achieving their objectives to grant medical assistance of quality. In this work, we will present the application of business processes modeling to the processes of an institution of the health sector, using the BPMN notation. The objective of this work is to show our experience obtained in the elaboration of the conceptual models of some hospital processes that can be used as a basis for others in the collaboration with hospitals for modeling their processes using BPMN. Hospital processes are very complex, and with their graphical visualization, their management and improvements are facilitated by means of the understanding and detection of possible failures.

1 INTRODUCTION

Business processes enable the activities of a company to be described in a manner which is understandable to all its users, with the purpose of analysis and design. Design means explicitly modeling, designing, simulating and redesigning the process as the organization learns what is possible. Due to the need to respond to competitive pressure or to business opportunities, business analysts need to restructure processes quickly (Smith *et al.*, 2002).

In business process modeling the main concept is that of the business processes themselves, which describe the activities involved in the business and how they relate to and interact with the necessary resources to achieve a goal for the organization. Some specific goals of business process modeling are: (Beck *et al.*, 2005; Erickson and Penker, 2000): 1. To ease the understanding of the key mechanisms of an existing business, 2. To serve as a basis for the creation of appropriate information systems that support the business, 3. To improve the current business structure and operation, 4. To show the structure of an innovated business, 5. To identify outsourcing opportunities and, 6. To facilitate the alignment of business specifications with the technical framework that IT development needs.

Similar to all types of organizations, in the health sector it is vitally important to keep its business processes up to date. This objective must not only be achieved through the continuous improvement of the services offered but also as a fundamental part of the quality programs in which it is immersed.

Business processes in hospitals are very complex and variable, due to the daily work which requires frequent reactions to the interim results of diagnostic processes as well as to unexpected medical instructions. When comparing the business processes in a hospital to the processes of other business areas, certain characteristic properties are evident, such as those pointed out by (Amberg and Gräber, 1996): a high number of cooperating organizational units, limited resources, a high ratio of manual activities, most medical processes can be

blueprinted only roughly in advance and details of medical processes are frequently changed.

From the point of view of computer science, the tendency of clinical practice is to move towards a shared care environment in which knowledge of clinical information systems (a combination of technology, data and people) should include definitions of all clinical processes' aspects, as well as the functions and responsibilities of the people involved in them (Colreavy, 2000). In this line, (Osterweil, 2006) argues that more precise and comprehensive process definitions are more effective bases for the kinds of definitive analyses that lead to successful improvement efforts in a more efficient manner.

Our interest is based on the conceptual modeling of the health-care sector processes by using the philosophy of business process models as a starting point. There are several important works dealing with this subject, such as those presented by (Framiñán *et al.*, 2004) and (Parra *et al.*, 2005a; Parra *et al.*, 2005b) in which business process modeling and simulation are applied to the health sector, particularly in processes such as telemedicine and hepatic post-transplant. On the other hand, (Graeber, 1997) had already carried out a similar work but on the basis of workflow management systems for the design of hospital information systems. In (Röhrig, 2002) an approach is presented through which existing business process descriptions are reused to analyze health care security requirements. Additionally, in the literature there are works related to business processes in the health sector, for example the presented by (Anyanwu *et al.*, 2003) as well as (Habing *et al.*, 2001).

In this work, we will present the application of the Business Process Modeling Notation (BPMN) for business process modeling in the health sector with the aim of showing our experience in the collaboration with the process modeling that can be applicable to any institution in the health sector.

Our intention when modeling hospital processes is that of facilitating the visualization and understanding of the activities that are carried out in the fulfilment of their mission. Another main target is to make visible the current processes (as-is models) for their analysis and comparison with the target processes (to-be models). This is one of the main problems that arise when clearly identifying the logical sequence of the real processes and their efficiency. This is due to the fact that, in most cases, hospital employees carry out their activities mechanically without having a theoretical basis.

This paper is organized as follows: In section 2, the BPMN notation for business process modeling will be detailed; in section 3 an overview of the collaboration context will be shown and in section 4 we will describe and illustrate our first experience in health-care process modeling. Next in section 5 we will present some lessons learned from this work. Finally, in section 6 some of the conclusions drawn from this work will be put forward.

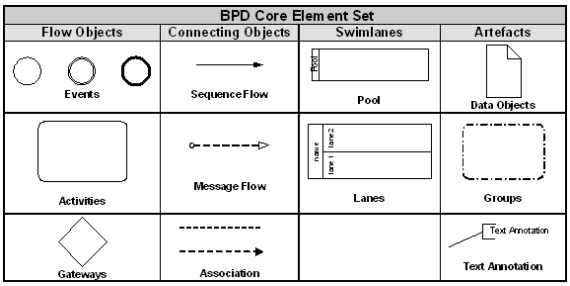
2 BUSINESS PROCESS MODELING WITH BPMN

Business process models (BPMs) can be created or presented by using many different techniques or languages. These languages are very different from one to another, since each one studies the processes in a different way, depending upon the purpose for which it was created (Dufresne and Martin, 2003). Among the languages for modeling business processes mentioned in the literature, special attention must be paid to the following ones: IDEF 0 (FIPS, 1993), IDEF 3 (Mayer *et al.*, 1995), UML (Erickson and Penker, 2000), UML 2.0 (OMG, 2003), and BPMN (OMG, 2006).

We are particularly interested in BPMN because its first goal is to provide a notation that can be easily understood by all business users, from the business analysts to the technical developers and business people (White, 2004). Moreover, it provides a graphical notation to express business processes in a Business Process Diagram (BPD), based on a flowcharting technique tailored to create graphical models of business process operations allowing the easy development of simple diagrams.

BPD is composed of two basic categories: the first one is formed by core elements *(Flow Objects*, *Connecting Objects*, *Swimlanes* and *Artifacts)* that make it possible to develop simple process models and a complete list of elements that allows the creation of complex or high-level business process models. Some of the BPMN elements for business process modeling are shown in Table 1.

Table 1: BPMN notation elements.



3 OVERVIEW OF THE COLLABORATION CONTEXT

With the objective of analyzing the use of BPMN notation in real cases, a multidisciplinary work group was created. The group was composed of software engineers from the Alarcos Research Group of the University of Castilla-La Mancha (UCLM) and health professionals and administrative staff from the General Hospital of Ciudad Real (GHCR) which is a part of the Health Care Services of the Spanish region of Castilla-La Mancha.

The main interest of the GHCR group is to build a state-of-the-art health care system, but it has faced several challenges in the application of business process modeling to health care processes. In this context, a set of work objectives and methodology to be followed were defined for each group. The applied research method was Action-Research (A-R). A-R is a collaborative research method aimed at joining theory and practice between researchers and practitioners by means of a process of a cyclical nature. A-R is focused on new knowledge building which must be useful in practice and which is obtained through searching for solutions to real situations (Avison *et al.*, 1999). Given its qualitative and cyclic nature, A-R is a very suitable method to apply to organizations in order to promote the continuous improvement of their processes. Being most advisable for our study, this one was the method that was selected.

On this basis, and in agreement with the research method, our first move was to approach the GHCR work team. The work in collaboration was carried out in different phases that are detailed below:

Phase 1: The first phase was started with three informative meetings that lasted three hours. In this first meeting with the hospital manager and the people in charge of the different areas of the hospital, the objectives of the process were put forward and a brief introduction to business process modeling and its advantages were presented; moreover team works were defined for choosing the processes to be modelled. Later on, two other meetings with the people assigned by the hospital team work to be responsible for the collaborative work took place.

Phase 2: In the second phase the personnel of the hospital received a 20 hours intensive course. This training session went into depth in the subject of modeling and the techniques for business processes modeling specifically BPMN notation, as well as in the importance of Business Process Management.

Phase 3: In this phase the hospital work group carried out meetings for the selection of three possible processes with different degree of complexity but equal importance level for the institution. This selection was mainly based on the following criteria:

- Processes must be well-known by all the participants in the work group.
- The process must affect multiple departments
- Information technology was considered as an essential tool during the workflow.
- Necessity of improvement of the process.

Phase 4: The next step was the definition of the work method by the members of this group. With the purpose of integrating the knowledge of the different members of the team, a specific work group was selected for each process, designing at least one representative member for each subprocess or affected organizational unit. Moreover, it was necessary to carry out the compilation of information, data and documents (such as textual definition of the process, paper-based documentation and electronic forms) that could be used for the elaboration of the selected process models.

Phase 5: In this phase, two meetings of both team works were carried out (GHCR and UCLM) that lasted around 2 hours each. In these meetings, the selected processes were: a) Programmed Surgical Patient (PSP), b) incorporation of a new employee, c) Citation process.

Phase 6: This phase consisted basically of the development of the PSP process model. The way of creation of the model was collaborative and iterative. In order to achieve this goal, there were necessary work meetings with the person in charge of the process for the refinement of the model to be carried out. Also interviews to some of the participant roles in the process execution took place.

4 MODELING A HOSPITAL PROCESS

In this work, we used the Business Process Modeling Notation (BPMN) for the health care sector processes modeling. We have selected this notation because it is widely accepted and recognized in the enterprise market due to the easiness that it provides for the construction of simple or high level processes. In this case, the model has to be as simple, transparent and understandable as possible for all the stakeholders in the health sector.

The process model presented in Figure 1 shows at a high level abstraction the Programmed Surgical Patient (PSP) process and the activities that are carried out when a patient is admitted into the hospital for the accomplishment of the surgical treatment of an injury or a disease that has been previously prescribed by means of diagnosis.

At the construction phase of this model a first rough draft was elaborated on the basis of the information that has been previously compiled (textual description of the process, documents, forms, etc.). Having elaborated a first version of the conceptual model of the process, a meeting with the stakeholder in charge of the process from the GHCR work group was carried out.

Therefore, simultaneously, the construction of the final model was possible, by means of the process description by the group of users to obtain a continuous feedback of the process. The construction of the model presented in Figure 1 was also possible thanks to the interviews made to the departmental personnel (participant user group and roles) involved in the process execution. By means of these interviews users explained to the modeler the work that they develop and the main activities that are carried out throughout this health care process. Once evaluated the PSP process model, the development of this model in a lower level of decomposition began. The first subprocess to develop is that corresponding to the Surgical Waiting List (SWL).

Similarly, for the construction of this subprocess the necessary textual and documentary information was compiled. Moreover, we have counted on the collaboration of the person in charge of this subprocess for its description and validation. The obtained result is shown in Figure 2.

With the development of this first model, and one of its subprocesses, it was possible to obtain the As-is model of one of the most representative processes within the sanitary service that offers the participant institution.

The work done will be the reference model that will serve as a basis for the construction of the "to-be"models. The application of the BPMN notation was very useful, since due to its characteristics; at any moment the construction of the processes models presented was facilitated for both participant groups (of business and systems).

5 LESSONS LEARNED

In the development of the "as-is" models of the process selected, there were some difficulties and limitations, mainly the following ones:

- Due to the nature of the work in the health sector, management in hospitals is oriented towards functions and not towards processes. This is because of the high degree of specialization of the different activities which frequently are made up of units with a high level of decentralization.

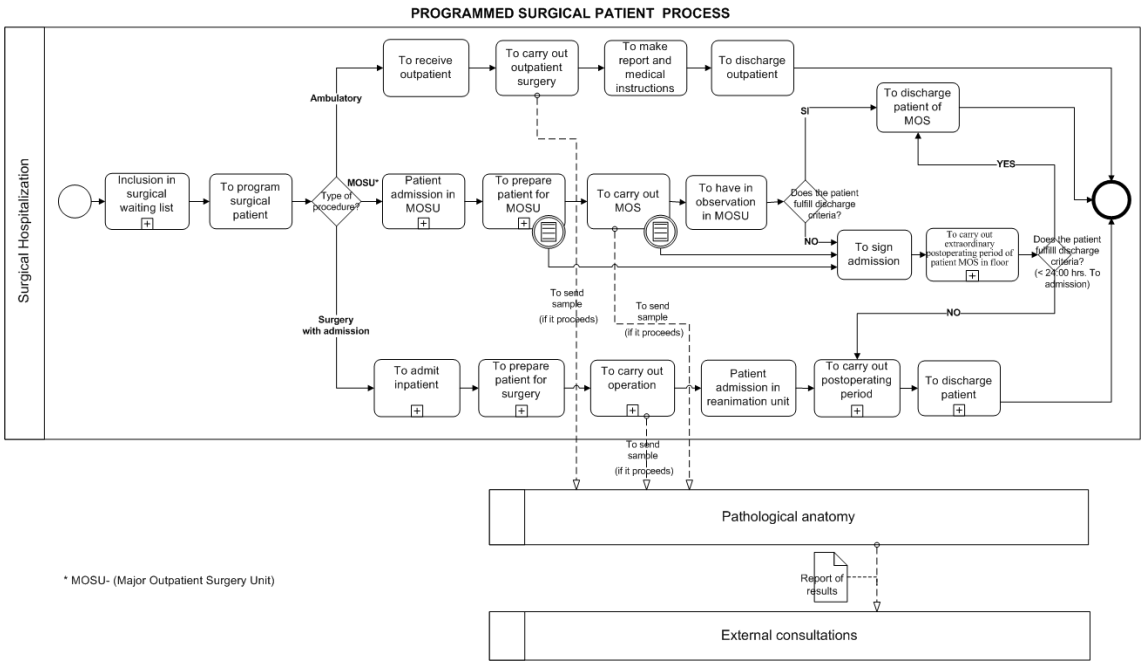


Figure 1: Programmed surgical patient process model.

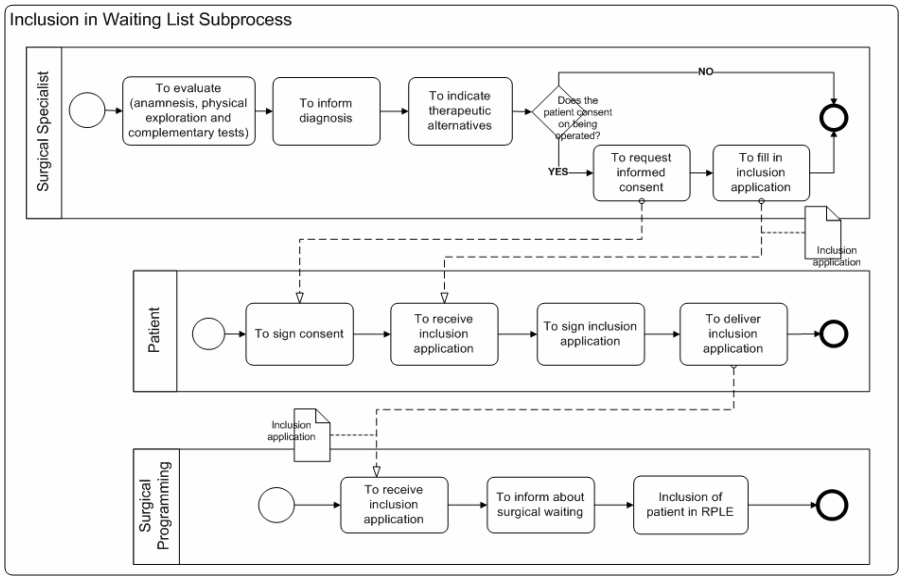


Figure 2: Surgical waiting list subprocess model.

- In the chosen process, diverse roles from different organizational departments which also work with certain degree of autonomy with respect to the others take part. This fact represented a problem at the time of efficiently designing the process.
- Variability of the process: Due to the fact that it is applicable to all surgical services of the hospital and to all patients who are going to take part in any surgical pathology except for urgent surgery, which entails many exceptions in decision making.
- There were some discrepancies between the participant professionals in the process at the time of the description of its activities, which partly reflects the absence of a model bases for the process.
- When the roles of the active people in charge in the process were not well defined a conflict was generated in the identification of the activities.
- In some cases the asynchronous communication by means of email was necessary.

Nevertheless, it is also important to emphasize the advantages that the application of business process modeling using BPMN in the health sector provides such as:

- The most important aspect was than with the training on BPMN that received the hospital staff, they had a clearer vision of their processes and the form in which they had to express it.
- Before creating the business process model, the textual description of processes had not a clear structure. After receiving the course, the personnel from the hospital could develop a more structured description of their processes in accordance with our objective consisting of modeling them by using BPMN.
- The process modeling serves as a reference as well as a starting point for the documentation of the processes that it carries out. In addition, the process modeling is able to be the basis for the certification of this process within the program of continuous improvement of the hospital.
- Through the graphical representation of the processes, the understandability of the model is facilitated being in addition a basis for decision making, as well as for redesigning future new processes.
- Having the models of the processes facilitates the incorporation and integration of new personnel into the working areas and processes of the hospital.
- Within the phase of analysis, some steps of the process were clarified such as that stating that members of the work group were different.
- It was possible to identify the different roles interacting between themselves within the process.
- Being a practically recently opened hospital it has inherited processes of the sanitary institutions that previously formed it. This is the reason why the design and modeling of its present processes come to represent an evolution approaching the quality of the service.
- With the detection of failures and bottle-necks in the present process will allow us to take corrective measures in the rules from this process, therefore, facilitating the management of its processes.

6 CONCLUSIONS AND FURTHER WORK

In this work we have presented an example of the application of business process modeling to the health sector. The main objective is in the first instance, to show on the basis of the accomplishment of this work the experience acquired in business process modeling applied to health sector processes and with the active participation of hospital staff (business people), and on the other hand to show the results obtained in the modelled and graphical visualization of some of their processes to starting off mainly by the received training. Once the first selected model was created, it was possible not only to determine the utility of process models for managing these processes but also to propose improvements of the same.

The experience in process modeling using the BPMN notation in the health sector has been interesting and fruitful for the work group. In addition, it was demonstrated that it is easily applicable and also, that it would be easily applicable in any other enterprise field and simultaneously understandable by all the users of the process.

This work and the lessons learned will be the starting point for the development of the models of those processes that the institution considers to be the most relevant ones in the fulfilment of its mission. Besides, this will allow the institution to have reference models that will be, among other things, the basis for: the certification within the program of continuous improvement of the hospital, the analysis and redesign of its processes, the evaluation of the efficiency of the process, the elaboration of hospital information systems, etc. In future works, we will complete the three processes chosen by the GHCR work group.

ACKNOWLEDGEMENTS

This work has been partially financed by the ENIGMAS Project (Junta de Comunidades de Castilla-La Mancha, Consejería de Educación y Ciencia, reference PBI-05-058), ESFINGE Project (Ministerio de Educación y Ciencia, Dirección General de Investigación/Fondos Europeos de Desarrollo Regional (FEDER), reference TIN2006-15175-C05-05).

REFERENCES

Amberg, M. and Gräber, S., 1996. Specifying Hospital Information Systems Using Business Process Modeling. Medical Informatics Europe. 1037-1041

Anyanwu, K., Sheth, A., et al., 2003. Healthcare Enterprise Process Development and Integration. Journal of Research and Practice in Information Technology 35(2): 83-98.

Avison, D., Lan, F., et al., 1999. Action Research. Communications of the ACM 42(1): 94-97.

Beck, K., Joseph, J., et al., 2005. Learn Business Process Modeling Basics for the Analyst, IBM.

Colreavy, P., 2000. Lack of adequate medical involvement in patient centred information technology. Stud. Health Technol. Inform. 77: 51-56.

Dufresne, T. and Martin, J., 2003. Process Modeling for E-Business. George Mason University.

Erickson, H.-E. and Penker, M., 2000. Business Modeling with UML- Business Patterns at Work.

FIPS, 1993. Integration Definition for Function Modeling (IDEF0). National Inst. of Standards and Technology.

Framiñán, J. M., Parra, C., et al., 2004. Experiencias en la aplicación de Modelado de Procesos de Negocio (BPM) en el sector sanitario. VIII Congreso de Ingeniería de Organización, Leganés. (Eds). 437-446

Graeber, S., 1997. The Impact of Workflow management Systems on the Design of Hospital Information Systems. AMIA Symposium.

Habing, N., Dietz, J., et al., 2001. Activity Patterns in Health Care - Identifying Building Blocks for the CPR. Workshop on the Language-Action Perspective on Communication Modelling, Montreal, 69-86

Mayer, R. J., Menzel, C. P., et al., 1995. Information Integration for Concurrent Engineering (IICE) IDEF3 Process Description Capture Method Report.

OMG, 2003. Unified Modeling Language (UML) Specification: Infrastructure, version 2.0.

OMG, 2006. Business Process Modeling Notation. Object Management Group. http://www.bpmn.org/

Osterweil, L. J., 2006. Ubiquitous Process Engineering: Applying Software Process Technology to Other Domains. Software Process Change. 3966: 39-47.

Parra, C., Framiñán, J., et al., 2005a. Modelado y Simulación de un Proceso de Atención Sanitaria: Aplicación al Proceso de Post-Transplante Hepático. Informática y Salud. 51: 25-35.

Parra, C., Framiñán, J., et al., 2005b. Metodología de Modelado de Procesos en Servicios de Salud utilizando Telemedicina. Inf. y Salud. 51:16-23

Röhrig, S., 2002. Using Process Models to Analyze Health Care Security Requirements. Conference Advance in Infrastructure for Electronic Business, Education, Science, Medicine and Mobile Technologies on the Internet,

Smith, H., Neal, D., et al., 2002. The Emergence of Business Process Management. CSC´s Research Services.

White, S. A., 2004. Introduction to BPMN. I. Corporation, bpmn.org (pub).

A NEW MODEL FOR SUCCESSFUL CPOE DEPLOYMENT AND IMPLEMENTATION IN HOSPITALS

Majid M. Altuwaijri
College of Public Health and Health Informatics
King Saud bin Abdulaziz University for Health Sciences (KSAU-HS), Saudi Arabia
majidt@ngha.med.sa

Keywords: CPOE, CSF, CPR, PMO, and Knowledge barriers.

Abstract: In spite of the importance of information technology (IT) for many health organizations to help manage the enterprise daily transaction, IT project failure rates still remain high. This suggests continued exploration of new process model and organization structure to nurture strong project performance. In this paper we propose a new model for successful implementation of IT projects. The proposed model calls for the establishment of a program management office (PMO) to implement corporate strategy for project management and to transform the organization into a learning one. The model is explained in details using an example of a Computerized Physician Order Entry (CPOE) implementation.

1 INTRODUCTION

Between 50,000 to 100,000 Americans die each year due to medical errors and about 1 million people are injured (IMO, 1999). In response to these chocking numbers reported by the Institute of Medicine (IOM), some healthcare organizations embarked on computerized physicians order entry (CPOE) systems.

Health care organization leaders are becoming well aware of the potential value of advanced clinical information systems such as Electronic Medical record (EMR) or computerized patient record (CPR) to improve outcomes, reduce medication errors, increase health care efficiency, and eliminate unnecessary costs. Many health organizations have invested significantly to plan, procure, and implement these advanced systems, including the current focus on computerized physician order entry (CPOE) (Haux, 2006). However, despite their knowledge, investments, and best intentions, most healthcare organizations have not realized a return on their investments (Berger, 2004). One probable reason for this problem is that the key users – physicians and other clinicians – are not using the technology to its greatest potential or, in too many cases, have not begun to use the technology at all. CPOE represents a monumental step forward for healthcare organizations because it embodies a shift from traditional, paper-based care coordination activities to automation of the order entry processes. This shift can be an agent for change, eliminating confusing or illegible hand-written order documentation, minimizing transcription errors and fundamentally reducing clinical mistakes (Snyder, 2006).

In this paper, a new model for successful implementation of CPOE project in hospitals. The model consists of three main components; the first component is concerned with the stages taken by organization to deploy new technology innovations which is in turned composed of four stages: visioning, matching vision, deployment, and evaluation and improvement. The second component deals with the main knowledge barriers to IT innovation diffusion as suggested by literatures. The third component is concerned with the critical success factors of IT innovation implementation as suggested by various literatures.

2 CPOE SYSTEM OVERVIEW

CPOE is a process of electronic entry of physician's orders and instructions for the treatment of patients. These orders are communicated over a CPR to the

medical staff (nurses, therapists or other physicians) or to the departments (pharmacy, laboratory or radiology) responsible for fulfilling/documenting the order.

CPOE is not a technology, rather it is a workflow design (or redesign) of clinical processes that integrates technology to optimize physician ordering of medications, laboratory tests, etc. (AHRQ, 2001). CPOE uses clinical decision support systems and links to hospital systems to generate prompts and alerts during the ordering session to notify of potential errors such as contra-indicated medications or routes or duplicate orders. Integration with other hospital information technology systems including electronic patient records, pharmacy, laboratory, and other services provides the physicians with all information necessary to develop and transmit an effective, error-free order (Metzger, 2003).

In May 2001, thirteen CPOE experts from around the world gathered at a 2-day conference for the purpose of developing recommendations for CPOE implementation (Ash, 2003). A list of high-level considerations was generated to benefit organizations thinking about implementing CPOE as follows:

- Motivation for Implementing POE
- Vision/Leadership/People
- Costs
- Integration/Workflow/Health Care Processes
- Value to Users/Decision Support Systems
- Technical Considerations
- Management of Project
- Training/Support/Help at the Elbow
- Learning/Evaluation/Improvement

3 PROJECT MANAGEMENT OFFICE

In this context, we adopt the PMO's definition proposed by PMI (PMI, 2004): "An organizational body or entity assigned various responsibilities related to the centralized and coordinated management of those projects under its domain. The responsibilities of the PMO can range from providing project management support functions to actually being responsible for the direct management of a project". A PMO focuses on the activities with relation to projects including training, process standardization, consulting, identifying of best practices, project prioritization, and reviewing project progress (Bolles, 2004).

4 INFORMATION SYSTEMS SUCCESS

DeLone and McLean in 1992 conducted a comprehensive review of IS success literature and proposed a model of IS success (Delone, 1992). It suggested that the success can be represented by the system quality, the output information quality, consumption (use) of the output, the user's response (user satisfaction), the effect of the IS on the behavior of the user (individual impact), and the effect of the IS on organizational performance (organizational impact). An updated model was proposed in 2003 (Delone, 2003) as shown in figure 1. The primary differences between the original and updated models included:

- The addition of Service quality category to reflect the importance of service and support in successful e-commerce systems,
- The addition of "intention to use" to measure user attitude, and
- The merging of individual impact and organization impact into one category concerned with net benefits.

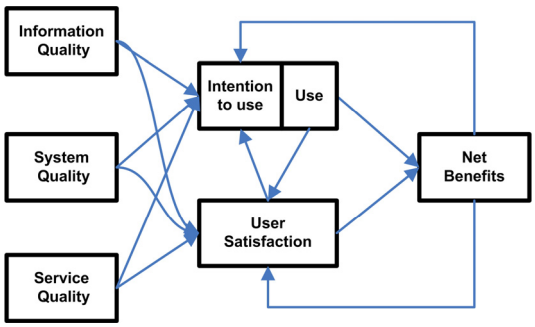


Figure 1: Updated Delone & McLean model of IS success.

5 KNOWLEDGE BARRIERS

Various study researches highlighted several barriers to innovation (Attewell, 1992; Tanriverdi, 1999; Nambisan, 1999). The adoption and implementation of complex IT solution is influenced by the organization's ability to lower or remove the various knowledge barriers. Knowledge barriers associated with the adoption of larger-scale IT solutions can be categorized into four categories: project/economic barrier, technical barriers, organization barriers, and behaviour barriers (Pare, 2007).

- Project/Economic barriers: This category is concerned with the financing and project management issues faced when acquiring innovation.
- Technological barriers: This category is concerned with the lack of knowledge required to carry out technical tasks needed to adopt new innovations.
- Organizational barriers: This category is concerned with the difficulties of deploying a new technology into existing practices and processes.
- Behaviour barriers: This group is concerned with the resistance to change among individuals affected by the implementation. It is also concerned with organizational power dynamics.

6 THE PROPOSED MODEL

Figure 2 shows a new model for successful implementation. The model consists of three main components; the first component is concerned with the stages taken by organization to deploy new technology innovations which is in turned composed of four stages: visioning, matching vision, deployment, and evaluation and improvement. In visioning phase, institutions define the corporate mission, objectives, and strategy. This phase is mainly concerned with identifying and prioritizing the organizational problems and opportunities that form the basis of the need to acquire innovations (Rogers, 1995). The second phase is concerned with the fit between a need identified in the first phase and the innovation proposed. In this stage we should determine whether the innovation will truly solve (or at least solve) one of the problems identified in the first phase. At the end of this phase, the organization decides weather or not to approve the innovation project.

In the event that the project is approved, the third phase of the process, deployment, begins. This phase includes all decisions and actions related to the deployment of the innovation. It includes also the assimilation and the integration of the innovation within the organization. At the end of this phase, the IT innovation solution is deployed within the organization. Evaluating performance is an important step for ensuring the quality of the innovation deployment. This phase emphasizes process flow optimization and continuous expansion of the system to gain competitive advantage.

The second component deals with the main knowledge barriers to IT innovation diffusion. The third component is concerned with the critical success factors of IT innovation deployment as suggested by various literatures (Ash, 2003).

After the successful deployment of innovation, the list of benefits and lessons learned feeds a knowledge base which in turn feeds all three components described in the model. Knowledge sharing behaviours facilitate learning among Project team members and enable them to resolve problems similar to situations encountered by others in the past, thus enabling more successful projects.

The proposed model calls for the establishment of a program management office to implement corporate strategy for project management. The main goal of this office is to translate the organization's strategic plan into projects and programs. The PMO is accountable for enterprise-wide distribution of project management best practices. Therefore, for the model to work properly, Organizations should invest in project management training for the staff working in the PMO.

The main advantages of this model is the strategic alignment of projects which bears on the synergy created by the management of relations between projects, and the ability to develop a better understanding of the challenges faced in carrying out information systems projects, the factors for success, and the strategies required to take advantage of IT. The acts of sharing are very important since a project's knowledge will not have much impact on the organization unless it is made available to other projects. Such learning organizations would be aware of the repeated knowledge barriers to innovation adoption and a well defined plan to address these barriers would be developed. Moreover, the knowledge base will help these organizations refine their strategies and prioritized plans. This allows for focusing on preparation for future projects which is rarely covered in literature.

7 APPLYING THE MODEL FOR CPOE DEPLOYMEENT

Figure 3 shows the detailed model for CPOE implementation as explained in the following sections:

7.1 Visioning

Institutions should align the strategic vision process with the budgeting process to produce a realistic prioritized strategic plan. The strategic plan will be used as an input to the IT strategic plan in order to ensure that the IT plan is closely integrated with the

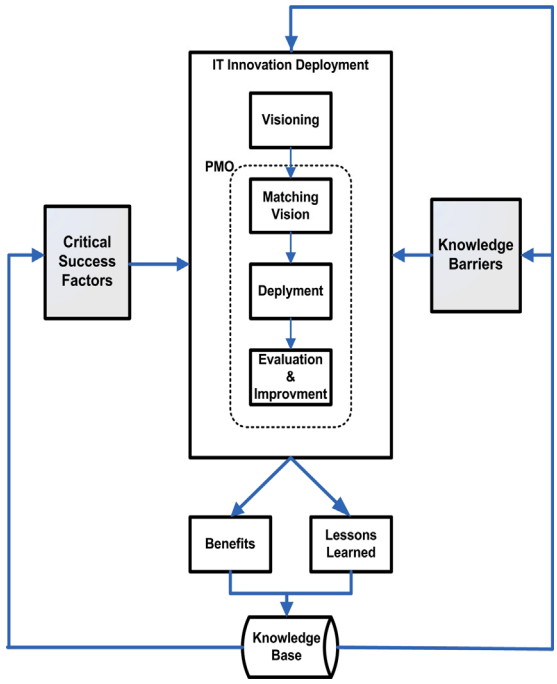


Figure 2: The proposed model for IT diffusion.

organization's strategic initiatives and business directions and provides the opportunity to use IT as a tool to enable systems integration as well as deliver information as a strategic resource. The visioning phase should address the Economic barrier. Moreover, there are some strategic CSFs related to this phase such as top management support, business plan and vision, and cost/benefit analysis.

7.1.1 Addressing the Economic/ Project Barrier

Organizations deploying CPOE solutions should address the financing issues faced when acquiring the solution. These barriers include barriers associated with project management skills. Moreover, multi-site implementation of CPOE presents special concerns. Firstly, each site had its own processes which may not be consistent with other sites. Secondly, multi-site organizations need to choose whether the implementation is done simultaneously in all facilities or in one facility at a time.

7.1.2 CSFs for the Visioning Phase

- Motivation for Implementation: The IT strategic plan along with the organization's strategies should provide a clear communicated business plan and vision to steer the direction of the CPOE project and other IT projects. This factor influences the funding, the political support, and the clinical leadership.
- CPOE Vision, Leadership, and Personnel: Successful deployment of IT innovation requires strong leadership, commitment, and continuous support by top management. A shared vision about the CPOE project should exist throughout the organization.
- Cost/Benefit Analysis: This analysis is instrumental for organizations in order to secure fund. This analysis would act as the basis for defining the success criteria for the CPOE project.

7.2 Matching the Vision

The PMO office plays an important role in matching the fit between projects, proposed by departments, and the organization vision and strategies. In this phase the organization should decide if the projects proposed match the vision or not. In case there is a match, a decision is made to invest on the proposed solution. If there is no match, the PMO office should update the organizational knowledge base and record the reasons. In some cases, the project proposals need further research before a decision is made. This phase is affected by the same barrier and CSFs as in the previous phase.

7.3 CPOE Deployment

7.3.1 CPOE Package Selection

Organization should make an extensive review of available CPOE Systems to select the package that has the best match of the organization's vision and strategies. The decision should be based on several factors including price, supplier support, ease of implementation, closeness to fit to organization’s business, technological risks, and local experience.

7.3.2 CPOE Project Life Cycle (PLC)

Organizations implementing CPOE systems usually modify business processes (as reasonably required) to conform to “best practices” discovered during the implementation of the CPOE project. These opportunities for process improvements would be identified during Business Process and Fit/Gap Analysis workshops, consulting visits and by the project team members as they become familiar and comfortable with the software. Process change

proposals would be considered and approved by the project management, steering and policy committees as part of the governance of the project. This phase is concluded when the CPOE solution is deployed and is running within the organization.

7.3.3 Addressing Knowledge Barriers

- Technical barriers: These include the technical infrastructure readiness. Organizations might need to upgrade their network infrastructure, upgrade the data centre's servers, and recruit technical and functional analysts to support the project. Data migration from legacy IT systems to the new CPOE is an essential task in this project. Finally, a maintenance contract is required to ensure the availability of the CPOE solution.
- Organizational barriers: These include change management processes to align the CPOE with existing practices and processes.
- Behaviour barriers: These include end-users resistance to new innovations. Organizations should conduct a lot of training sessions to introduce the new innovation.

7.3.4 Critical Success Factors

- Integration: Organization should have a clear plan for integrating the CPOE into existing environments and workflows. This factor is also concerned with integrating the CPOE into other hospital applications and possible legacy systems.
- Value to Users: Organizations should have a clear change management strategy which includes analysis of the current business processes to identify the areas for improvement. Physicians must be shown CPOE benefits including the embedded decision support logic that helps to improve the quality of patient care.
- Project management: Organizations should appoint a dedicated full-time project manager. The management of the project covers the technical and functional aspects. A qualified project manager manages five main functions; managing scope, managing project team, managing quality, managing cost, and managing time.
- Technology: Organizations should make sure they have a good technical infrastructure before "Go-Live". The project team should make a balance between customization and standardization; where physicians of the same specialty would have a common CPOE screen which might be different than screens for different specialties. Organizations should also have a comprehensive system testing plan to test the functionality of each module alone and in conjunction with other modules.
- Training and support: Organizations should have a comprehensive training plan which includes functional and technical aspects of the project. In large organizations, "train the trainer" approach is usually considered. Additionally, Organizations should consider proper support plan including 24*7 support for at least several days post go-live.

7.4 Evaluation and Improvement

Finally, evaluation and improvement phase come into action after the CPOE system is already deployed within an organization. It is important that organizations define the success criteria of the CPOE project and use it to monitor the project. CPOE system should be monitored forever and should receive continuous improvement. Organizations usually assume that their operational performance will be improved immediately after the "go-live". Instead, due to the complexity of CPOE systems, organizations should expect an initial decline in their performance. Once the system is stable and users are more aware of its capabilities, improvements will take place.

8 CPOE BENEFITS, LESSONS LEARNED

After the CPOE system is already deployed, the PMO should measure the benefits realized by the project and tie them to the organizational strategic goals. It is also important that the PMO assembles a list of lessons learned from this project. The list of project benefits and lessons learned are then fed into the organization knowledge base. The knowledge base will provide projects information back to the organization visioning process and will be used to refine the list of CSFs and knowledge barriers for future projects. The feed back loop will transform the organization to becoming a learning one.

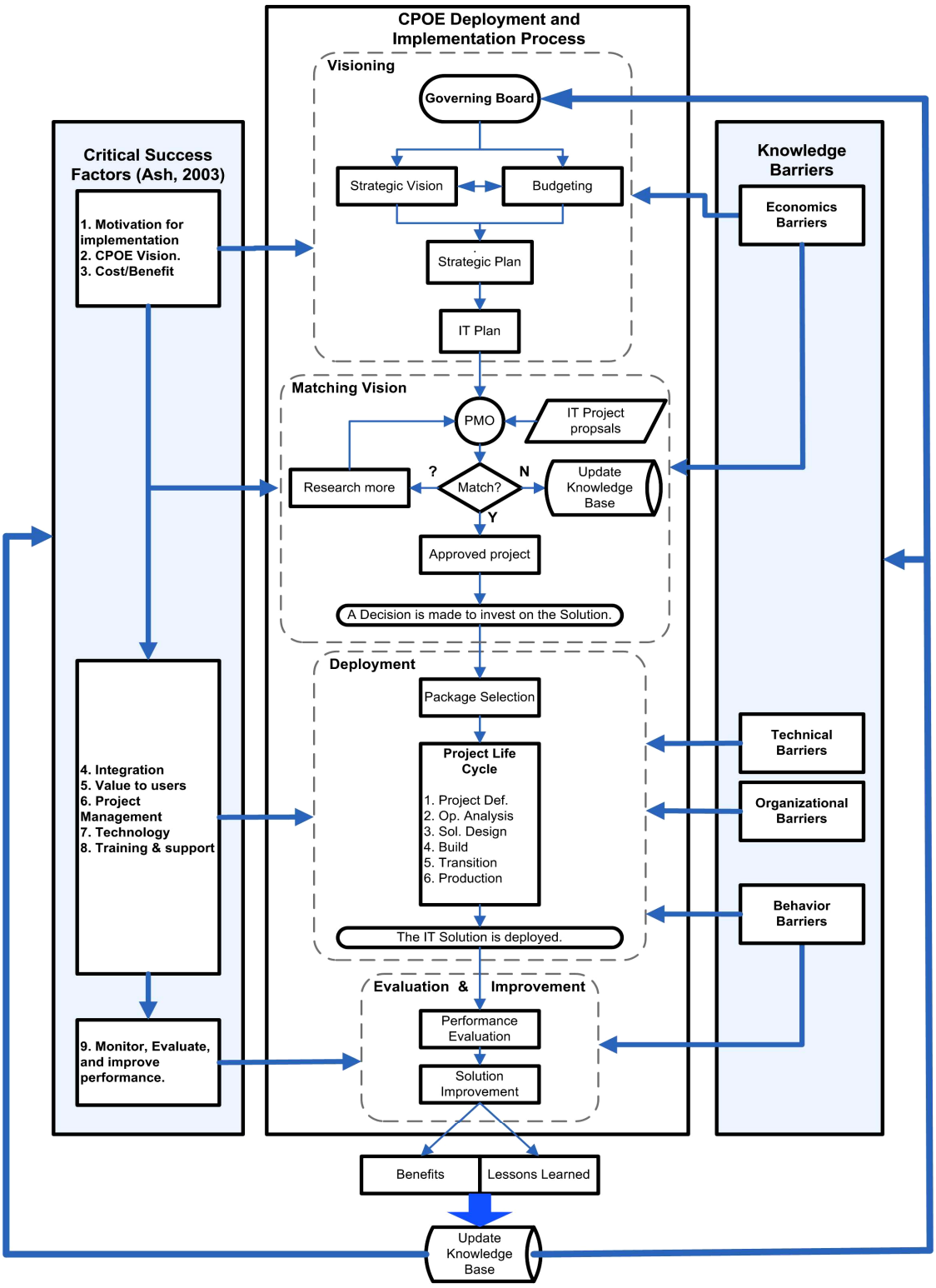


Figure 3: The proposed model for CPOE deployment and implementation in hospitals.

9 CONCLUSIONS

Enterprise-wide daily transactions are difficult to manage due to their complexity and scope in terms of rendering services to their clients. As a result, many organizations employ Information Technology (IT) to manage their enterprise-wide transactions and to ensure quality of services. IT, when properly selected and implemented, helps organizations increase their efficiency and capabilities which improves the organization's competitive advantages. In spite of the importance of information technology (IT), IT project failure rates still remain high. This suggests continued exploration of new process model and organization structure to nurture strong project performance.

In this paper we propose a new model for successful implementation of IT projects. The model consists of three main components; the first component is concerned with the stages taken by organization to deploy new technology innovations which is in turned composed of four stages: visioning, matching vision, deployment, and evaluation and improvement. The second component deals with the main knowledge barriers to IT innovation diffusion as suggested by literatures. The third component is concerned with the critical success factors of IT innovation implementation as suggested by various literatures. The proposed model calls for the establishment of a program management office to implement corporate strategy for project management and to transform the organization into a learning one. The model is explained in details thru an example of CPOE implementation.

The acts of sharing are very important since a project's knowledge will not have much impact on the organization unless it is made available to other projects. Although knowledge exists at different levels of an organization, for instance, at the individual, team, and organization levels, sharing of knowledge at the individual level is critical to an organization.

It is expected as organizations use this model for several projects; more knowledge would be gained and used towards more successful project in the future. Critical success factors would be refined more and be closely tied with organization culture. Such learning organizations would be aware of the repeated knowledge barriers to innovation adoption and a well defined plan to address these barriers would be developed. Moreover, the knowledge base will help these organizations refine their strategies and prioritized plans.

REFERENCES

Institute of Medicine (1999). To Err is Human: Building a Safer Health System. *National Academy Press*, Washington, DC.

Haux, R. (2006). Health Information systems-past, present, future. *Int. J. of Med. Info., 75, 268-281.*

Berger, R. G. & Kichak, J. P. (2004). Computerized Physicians Order Entry: Helpful or Harmful. *J. Am. Med. Assoc., 11*, 100-103.

Snyder, R., Weston, M., Fields, W., Rizos, A. & Tedeschi, C. (2006). Computerized provider order entry system field research: The impact of contextual factors on study implementation. *Int. J. of Med. Info., 75*, 730-740.

AHRQ (2001). *Making Health Care Safer: A Critical Analysis of Patient Safety Practices*. Found in http://www.ahrq.gov/clinic/ptsafety/

Metzger, J. & Fortin, J. (2003). *Computerized Physician Order Entry in Community Hospitals: Lessons from the Field.* California HealthCare Foundation and First Consulting Group.

Ash, J. S., Savri, P. Z. & Kuperman, G. J. (2003) .A Consensus on Considerations for a Successful CPOE Implementation. *J. Am. Med. Assoc., 10,* 229-234.

Project Management Institute. *A guide to the project management body of knowledge 3rd edition.* Newton square, OA, 2004.

Bolles, D. (2004) Building project management centers of excellence. *AMACOM.*

DeLone, W. H. & McLean, E. R. (1992). Information systems success: the quest for the dependent variable. *Information Systems Research, 3(1)*, 1992, 60–95.

DeLone, W. H. & McLean, E. R. (2003). The DeLone and McLean model of information system success: a ten-year update. *Journal of Management Information Systems, 19(4),* 9–30.

Attewell, P. (1992). Technology diffusion and organizational learning: the case of business computing. *Organ, Sci. 3*, 1-19.

Tanriverdi, H. & Iacono, C. S. (1999). Knowledge barriers to diffusion of telemedicine. *in proceedings (ICIS 20th)*, Charlotte, NC, 39-50.

Nambisan, S. & Wang, Y.-M (1999). Roadblocks to web technology adoption? *in Commun. ACM. Vol. 42(1),* 98-101.

Pare, G. & Trudel, M.-C.(2007). Knowledge barriers to PACS adoption and implementation in hospitals. *Int. J. of Med. Informatics, 76,* 22-33.

Rogers, E. M. (1995). *Diffusion of Innovations, 4th edition*, Free press, New York, NY.

COMPARISON OF THREE NEURAL NETWORK CLASSIFIERS FOR APHASIC AND NON-APHASIC NAMING DATA

Antti Järvelin
Department of Computer Sciences, University of Tampere, FIN-33014 University of Tampere, Finland
antti.jarvelin@cs.uta.fi

Keywords: Neural networks, classification, aphasia, anomia, naming disorders.

Abstract: This paper reports a comparison of three neural network models (Multi-Layer Perceptrons, Probabilistic Neural Networks, Self-Organizing Maps) for classifying naming data of aphasic and non-aphasic speakers. The neural network classifiers were tested with the artificial naming data generated from confrontation naming data of 23 aphasic patients and one averaged control subjet. The results show that one node MLP neural network performed best in the classification task, while the two other classifiers performed typically 1 - 2 % worse than the MLP classifier. Although the differences between the different classifier types were small, these results suggests that a simple one node MLP classifier should be preferred over more complex neural network classifiers when classifying naming data of aphasic and non-aphasic speakers.

1 INTRODUCTION

Aphasia is a language impairment following left hemisphere damage. Aphasic patients have defect or loss of production or receptive aspects of written or spoken language (Harley, 2001). The most common symptom of aphasia is anomia, the impairment in word retrieval, which has devastating effects on patients ability to carry on meaningful and effective conversation (Raymer and Rothi, 2002).

The language capabilities of the aphasic patients are tested with standardized aphasia examination procedures, such as Boston Diagnostic Aphasia Examination (Goodglass and Kaplan, 1983) (in English speaking countries) or Aachen Aphasia Test (Huber et al., 1984) (in German speaking countries). An integral part of these tests is a picture confrontation naming task, where a subject is to name (i.e., say aloud) single pictures. The picture naming task is used, because picture naming process involves all the major processing stages of speech production (Laine et al., 1992). Thus, picture naming task may more clearly reveal the underlying mechanism and nature of patient's anomia than the plain analysis of free speech would (Dell et al., 1997; Cuetos et al., 2002).

Examples of common error types encountered in the naming test include semantic errors ("cat" → "dog"), formal errors ("cat" → "mat"), nonword errors ("cat" → "tat"), mixed errors, where the response is semantically and phonologically related to the target ("cat" → "rat") and finally unrelated word errors where no semantic or phonological relationship can be found between the target and produced word.

The goal of the current study was to investigate the suitability of neural network classifiers for separating healthy individuals from aphasic patients. Three neural network classifiers, Multi-Layer Perceptrons (MLP) (Haykin, 1999), Probabilistic Neural Networks (PNN) (Specht, 1990), and Self-Organizing Maps (SOM) (Kohonen, 1998), were compared for the classification of aphasic and non-aphasic naming data. The performance of the classifiers were compared using the aphasic naming data reported by Dell et al. (1997) which was artificially augmented to suit better for the neural network classifiers.

To our knowledge this kind of research has not been previously reported. For example, Axer et al. (2000) used a MLP classifier and *k*-nearest neighbor classifier recognize patients' aphasia type with their quite large data set. However, their data set did not contain control data from healthy subjects and thus it cannot be used in the current study.

2 MATERIALS AND METHODS

2.1 Materials

To compare the classifiers, aphasic naming data from Dell et al. (1997) was used (see Table 1). The data set contains six attributes describing the Philadelphia Naming Test (PNT) (Roach et al., 1996) results of the subjets: correct answers, semantic errors, formal errors, nonword errors, mixed errors and unrelated word errors. The original data also contained a category for all other miscellaneous errors, but it was excluded from the current study as redundant. The exclusion however explains why the error distributions of the patients rarely sum up to 1.

The data set of Dell et al. is quite small as it only contains naming performances of 23 aphasic patients and an averaged naming distribution of 60 healthy control subjects. Therefore, the data set was augmented to be able to use the neural network classifiers successfully for the task.

The data was augmented with the following procedure using the naming distributions of Table 1 as a basis of data generation.

1. First 23 artificial control subjects were generated from the averaged control subject and then the generated control subjects and the 23 original patient cases were merged producing the base set of 46 subjects.
2. From the base set 10×10 cross-validation sets were prepared.
3. Finally, the partitions of the cross-validation sets were augmented so that the total size of the cross-validation sets were 1000 cases. Thus, each patient and generated control subject served approximately 22 times as a seed for data generation.

The first part of the data generation ensures an equal class distribution between the healthy and patient data. It also ensures that there will be more variation in the generated healthy data than there would be if only the average control subject had been used as the seed for all generated healthy data. The second and third parts ensure that during cross-validation the test and validation sets always contain cases generated from the different seeds than the cases in the training set making the cross-validation process more robust.

The values of the variables for each generated test case were calculated as follows. Let v_i be the value of the ith variable of the seed, $\sigma(v_i)$ the standard deviation of variable v_i and $N(a,b)$ normal distribution with mean a and standard deviation b. The value v_i' of the ith variable of the artificial subject was calculated with

$$v_i' = |v_i + N(0, 0.1\sigma(v_i))|. \quad (1)$$

Applying (1) in the data generation produces a cloud of artificial subjects centered around the seed. The absolute value is taken in (1) in order to avoid negative values for the generated variables. Different scaling factors of the variables' standard deviations were experimented, and 0.1 was found to be the most appropriate one. A greater scaling factor would have dispersed the generated cloud around the center too much and the smaller would have resulted too compact clouds.

2.2 Methods

The neural network classifiers were compared generating ten data sets with the procedure described in the previous section. Ten data sets were used to smooth the differences between the generated data sets. Each classifier was examined by running a 10×10 cross-validation (Duda et al., 2001) for each data set. The differences between the classifiers were compared by calculating average classification accuracy (ACC) for each classifier over the ten cross-validated data sets. The total classification accuracy for a classifier is given by

$$ACC = 100 \cdot \frac{\sum_{c=1}^{C} tp_c}{\sum_{c=1}^{C} p_c}\%, \quad (2)$$

where tp_c denotes the number of true positive classifications and p_c the size of the class c. For more detailed evaluation also true positive rates (TPR) and positive predictive values (PPV) were calculated for the both classes. The true positive rate for class c is given by

$$TPR_c = 100 \cdot \frac{tp_c}{p_c}\%, \quad (3)$$

and the positive predictive value with

$$PPV_c = 100 \cdot \frac{tp_c}{tp_c + fp_c}\%, \quad (4)$$

where fp_c denotes the number of false positive classifications of the class c. The statistical significance of the differences between the classifiers was tested with Friedman test (Connover, 1999) over the classification accuracies.

For MLP, sigmoid activation function was used as network's transfer function. The networks were trained using Levenberg-Marquardt optimized batch mode backpropagation algorithm. To prevent over fitting a separate validation set (chosen from the cross-validation set) was used for early stopping during the training. The networks were trained at most 100 epochs, but typically less than 100 epochs were used, since the validation set brought the algorithm into

Table 1: The proportional naming distributions of 23 aphasic patients and an averaged control subject tested with PNT reported by Dell et al. (1997).

	Naming response					
Patient	Correct	Semantic	Formal	Nonword	Mixed	Unrelated
W.B.	0.940	0.020	0.010	0.010	0.010	0.000
T.T.	0.930	0.010	0.010	0.000	0.020	0.000
J.Fr.	0.920	0.010	0.010	0.020	0.020	0.000
V.C.	0.870	0.020	0.010	0.030	0.010	0.000
L.B.	0.820	0.040	0.020	0.090	0.010	0.010
J.B.	0.760	0.060	0.010	0.050	0.020	0.010
J.L.	0.760	0.030	0.010	0.060	0.030	0.010
G.S.	0.700	0.020	0.060	0.150	0.010	0.020
L.H.	0.690	0.030	0.070	0.150	0.010	0.020
J.G.	0.550	0.060	0.080	0.180	0.040	0.030
E.G.	0.930	0.030	0.000	0.010	0.020	0.000
B.Me.	0.840	0.030	0.010	0.000	0.050	0.010
B.Mi.	0.830	0.050	0.010	0.010	0.020	0.010
J.A.	0.780	0.040	0.000	0.020	0.030	0.010
A.F.	0.750	0.020	0.030	0.070	0.060	0.040
N.C.	0.750	0.030	0.070	0.080	0.010	0.000
I.G.	0.680	0.090	0.050	0.020	0.030	0.010
H.B.	0.610	0.060	0.130	0.180	0.020	0.010
J.F.	0.560	0.140	0.010	0.020	0.110	0.010
G.B.	0.390	0.070	0.090	0.080	0.010	0.030
V.P.	0.280	0.070	0.110	0.040	0.050	0.170
G.L.	0.280	0.040	0.210	0.300	0.030	0.090
W.R.	0.080	0.060	0.150	0.280	0.050	0.330
Control	0.969	0.012	0.001	0.000	0.009	0.003

early stop after few dozens of epochs. Totally six different network architectures were tested (1, 2-1, 3-1, 4-1, 5-1, 6-1, where *x*-*y* corresponds to the number of hidden nodes (*x*) and output nodes (*y*)).

With PNNs, the standard learning algorithm was used. In PNN learning algorithm, the only parameter that needs to be specified by the user is the width of the Gaussian window determining the radius of the activation functions in the network. Six different Gaussian window widths were experimented (0.01, 0.02, 0.03, 0.04, 0.05 and 0.06).

For SOM, the standard SOM algorithm was used. The network was trained totally 10000 iterations (ten epochs of the training data) of which 1000 iterations were used for initial ordering phase (with learning rate 0.9) and the rest for the convergence phase (with learning rate 0.02). After teaching, the class labels were assigned for each node using majority labeling (see e.g. (Kohonen, 2001)). Eight different SOM lattice architectures were tested (1×4, 1×5, 1×6, 2×2, 3×3, 4×4 and 5×5).

3 RESULTS

The results for each classifier are presented in Table 2. Based on the total classification accuracy, the MLP architecture seems to be the best choice from the tested neural network types for the classification task. The best performing MLP had average accuracy over 2 % higher than the best performing SOM network and over 1 % higher than the best performing PNN network. Also the standard deviations of the classification accuracies followed the same order, with MLP having the smallest deviation.

The TPRs show that the healthy class was easier for the classifiers to recognize than the patient class. MLP performed worst on the healthy class but best on the patient class, whereas SOM performed best with the healthy class but worst with the patient class. The differences of the TPRs between the classes were high for all classifiers. The classifiers also recognized the healthy cases almost perfectly, but the patient cases were harder to recognize. The differences of the TPRs between the classes were well over 10 % with PNN

Table 2: Means and standard deviations (%) of true positive rates (TPR), positive predictive values (PPV) and total classification accuracies (ACC) for the three best architectures of each tested neural network classifier type. The best architecture for each classifier type is in bold.

Classifier		TPR		PPV		ACC
Type	Architecture	Healthy	Patient	Healthy	Patient	
MLP	6-1	95.7 ($\pm$ 15.6)	86.4 ($\pm$ 22.9)	88.2 ($\pm$ 19.8)	94.2 ($\pm$ 17.8)	91.2 ($\pm$ 13.1)
MLP	5-1	94.8 ($\pm$ 18.2)	87.7 ($\pm$ 21.4)	88.2 ($\pm$ 21.1)	95.0 ($\pm$ 15.5)	91.4 ($\pm$ 13.6)
MLP	**1**	**97.5** ($\pm$ **5.3**)	**88.9** ($\pm$ **20.3**)	**92.0** ($\pm$ **13.0**)	**96.6** ($\pm$ **10.0**)	**93.3** ($\pm$ **9.7**)
PNN	0.03	99.5 ($\pm$ 1.8)	81.8 ($\pm$ 23.8)	87.1 ($\pm$ 16.0)	99.2 ($\pm$ 3.8)	90.6 ($\pm$ 11.9)
PNN	0.02	99.2 ($\pm$ 2.6)	82.7 ($\pm$ 23.2)	87.7 ($\pm$ 15.8)	99.1 ($\pm$ 2.8)	90.9 ($\pm$ 11.6)
PNN	**0.01**	**99.2** ($\pm$ **2.1**)	**84.4** ($\pm$ **21.9**)	**88.8** ($\pm$ **14.7**)	**98.9** ($\pm$ **4.9**)	**91.8** ($\pm$ **10.8**)
SOM	1×7	94.9 ($\pm$ 7.5)	86.1 ($\pm$ 19.8)	89.4 ($\pm$ 14.3)	94.9 ($\pm$ 6.9)	90.5 ($\pm$ 9.9)
SOM	1×4	95.4 ($\pm$ 7.3)	86.1 ($\pm$ 19.9)	89.5 ($\pm$ 14.3)	95.4 ($\pm$ 6.4)	90.8 ($\pm$ 9.9)
SOM	$\mathbf{1 \times 6}$	**99.5** ($\pm$ **1.6**)	**82.2** ($\pm$ **24.2**)	**87.6** ($\pm$ **15.8**)	**98.6** ($\pm$ **9.5**)	**90.9** ($\pm$ **11.9**)

and SOM and almost 10 % with MLP. The difficulty of recognizing the patient class is reflected also with the standard deviations of the TPRs of the patient class which were considerably higher than those of the healthy class.

The PPVs show that the most of the misclassifications for all classifier types occurred when a patient was classified into the healthy class. Indeed, the results show that confidence for the correct decision is high for all classifiers when they decided that a sample belongs to the patient class. For the best MLP classifier the PPVs of the both classes were well over 90 % and the difference between the two classes was only 4.6 %. Again, for PNN and SOM the PPVs for the patient class were extremely high, but the overall performance was deteriorated with significantly lower PPV values for the healthy class. The differences between PPVs of the classes were over 10 % for the best PNN and SOM classifiers.

Based on the Friedman test results, MLP and PNN performed statistically equally well. However, the differences between the classification accuracies of MLP and SOM and PNN and SOM were statistically significant with SOM being statistically an inferior classifier than the others. These results were statistically highly significant (at level $\alpha = 0.001$).

For MLP, the best performing architecture was a network with only one node. It generally had classification accuracy almost 2 % higher than the other tested architectures. PNN performed the best with Gaussian window width of 0.01. Generally, increasing the window width deteriorated the performance of the network and the difference between the accuracies of a network with the best performing window width 0.01 and the worst performing window width 0.06 was ca 2 %. For SOM, the best performing lattice structure was 1×6. Other tested lattice structures performed slightly worse their classification accuracies being 1 – 2 % lower than the best performing 1 $\times$ 6 lattice structure.

4 CONCLUSIONS

The results show clearly that MLP performed best in separating the aphasic speakers from healthy speakers based on their naming data distributions. MLP's total classification accuracy was 1 – 2 % higher than the accuracies of the other classifier types, and their smaller standard deviations of the classification accuracies proved it also to be a more robust classifier than the other tested classifiers. Furthermore, because only one node was needed to implement the most successful MLP architecture, other simple classification methods, such as discriminant analysis and Bayes classifiers should also perform well at the classification task.

Based on the TPRs and PPVs all classifiers were biased towards the healthy class. The MLP architecture favors least healthy data over patient data resulting into highest overall classification accuracy of the compared classifiers. The other two classifiers were even more biased towards the healthy class and correspondingly their classification accuracies were slightly lower than MLP classifiers. The SOM classifier was clearly the weakest classifier type and the differences of the classification accuracies to the other two classifiers were statistically significant. Based on these results, SOM classifiers should not be preferred in patient / healthy classification over simpler classification methods.

The differences between the average classification accuracies of the classifiers were very small and all classifiers had total classification accuracy over 90 %. Therefore, it seems that the choice of classification

method is not very crucial for the reported data set. This also supports the choice of the simplest classifier type, the one node MLP, as a preferred classifier. However, it has to be noted that the classification task might have been harder if more real data had been available serving as a basis for the data generation.

On the other hand, the results with the current data suggest that neurolinguistic tests used in aphasia testing separate quite well the healthy subjects and patients from each other. Thus, it is possible that there is no need for using more advanced classification methods in patient / healthy subject separation.

The current classification research should be extended to include more classifier types. Especially traditional classifiers types, such as naïve Bayes classifier, discriminant analysis or *k*-nearest neighbor classifiers should be investigated, since the success of the one neuron MLP classifiers suggest that the simple classification methods might perform noticeably well with the classification task.

An important question is also how well the used data generation method preserves the characteristics of the original data. This question should be examined in more detail in order to ensure that the data generation does not unnaturally bias the data. Moreover, other aphasia data sets should be tested, although finding a suitable data set with decent number of test cases seems to be problematic.

ACKNOWLEDGEMENTS

The author gratefully acknowledges Professor Martti Juhola, Ph.D., from Department of Computer Sciences, University of Tampere, Finland for his helpful comments on the draft of the paper and Tampere Graduate School in Information Science and Engineering (TISE) for financial support.

REFERENCES

Axer, H., Jantzen, J., and von Keyserlingk, D. G. (2000). An aphasia database on the Internet: a model for computer-assisted analysis in aphasiology. *Brain and Language*, 75(3):390–398.

Connover, W. J. (1999). *Practical Nonparametric Statistics*. John Wiley & Sons, New York, NY, USA, 3 edition.

Cuetos, F., Aguado, G., Izura, C., and Ellis, A. W. (2002). Aphasic naming in spanish: predictors and errors. *Brain and Language*, 82(3):344–365.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., and Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, 104(4):801–838.

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern Classification*. John Wiley & Sons, New York, NY, USA, 2 edition.

Goodglass, H. and Kaplan, E. (1983). *The Assessment of Aphasia and Related Disorders*. Lea & Febiger, Philadelphia, PA, USA, 2 edition.

Harley, T. (2001). *The Psychology of Language*. Psychology Press, New York, NY, USA, 2 edition.

Haykin, S. (1999). *Neural Networks. A Comprehensive Foundation*. Prentice Hall, London, United Kingdom, 2 edition.

Huber, W., Poeck, K., and Weniger, D. (1984). The Aachen aphasia test. In Rose, F. C., editor, *Advances in Neurology. Progress in Aphasiology*, volume 42, pages 291–303. Raven, New York, NY, USA.

Kohonen, T. (1998). The self-organizing map. *Neurocomputing*, 21(1–3):1–6.

Kohonen, T. (2001). *Self-Organizing Maps*. Springer-Verlag, Berlin, Germany, 3 edition.

Laine, M., Kujala, P., Niemi, J., and Uusipaikka, E. (1992). On the nature of naming difficulties in aphasia. *Cortex*, 28:537–554.

Raymer, A. M. and Rothi, L. J. G. (2002). Clinical diagnosis and treatment of naming disorders. In Hillis, A. E., editor, *The Handbook of Adult Language Disorders*, pages 163–182. Psychology Press, New York, NY, USA.

Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., and Brecher, A. (1996). The Philadelphia Naming Test: Scoring and rationale. *Clinical Aphasiology*, 24:121–133.

Specht, D. (1990). Probabilistic neural networks. *Neural Networks*, 3(1):109–118.

ARTIFICIAL NEURAL NETWORKS FOR DIAGNOSES OF DYSFUNCTIONS IN UROLOGY

[1]David Gil Mendez, [2]Magnus Johnsson, [1]Antonio Soriano Paya and [1]Daniel Ruiz Fernandez
[1]*Computing Technology and Data Processing, University of Alicante, Spain*
{*dgil, soriano, druiz*}*@dtic.ua.es*
[2]*Lund University Cognitive Science & Dept. of Computer Science, Lund University, Lund, Sweden*
Magnus.Johnsson@lucs.lu.se

Keywords: Artificial neural networks, urology, artificial intelligence, medical diagnosis, decision support systems.

Abstract: In this article we evaluate the work out of artificial neural networks as tools for helping and support in the medical diagnosis. In particular we compare the usability of one supervised and two unsupervised neural network architectures for medical diagnoses of lower urinary tract dysfunctions. The purpose is to develop a system that aid urologists in obtaining diagnoses, which will yield improved diagnostic accuracy and lower medical treatment costs. The clinical study has been carried out using the medical registers of patients with dysfunctions in the lower urinary tract. The current system is able to distinguish and classify dysfunctions as areflexia, hyperreflexia, obstruction of the lower urinary tract and patients free from dysfunction.

1 INTRODUCTION

Nowadays, the urologists have different tools available to obtain urodynamical data. However, it still remains very complicated to make a correct diagnosis: the knowledge concerning the origin of the detected dysfunctions depends mainly on acquired experience and on the research, which is constantly carried out within the field of urology. The specialists in urology are quite often dealing with situations that are poorly described or that are not described in the medical literature. In addition there are numerous dysfunctions whose precise diagnoses are complicated. This is a consequence of the interaction with the neurological system and the limited knowledge available on how this operates.

Various techniques are used to diagnose dysfunctions of the lower urinary tract (LUT), which entail different degrees of invasiveness for the patient (Abrams, 2005). A urological study of a patient consists of carrying out various costly neurological as well as physical tests like flowmetry and cystometry examinations with high degrees of complexity and invasiveness. This project is intended to aid the specialist in obtaining a reliable diagnosis with the smallest possible number of tests. To this end we propose the use of artificial neural networks (ANNs) since these present good results for classification problems (Begg, 2006). The reason why we decided to apply ANNs instead of other artificial intelligence (AI) methods for the support of medical diagnosis is due to the fact that ANNs can be trained with appropriate data learning in order to improve their knowledge of the system. In comparison to other techniques or other (more classical) approaches such as rules based systems or probabilities, ANNs are more suitable for medical diagnosis. On one hand, rules based systems (MYCIN system (Mycyn, 1976)) contain ”if-then” rules where the ”if” side of any rule is a collection of one or more conditions connected by logical operators such as ”AND”, ”OR” and ”NOT”. On the other hand, other systems such as probability systems calculate measures of confidence without the theoretical underpinnings of probability theory. These formal approaches based on probability theories are precise but can be awkward and non-intuitive to use.
Therefore, in medical diagnosis, we use the advantages of ANNs, which are considered as a method of disease classification. This classification has two divergent meanings. We can have a set of registers, vectors or data with the aim of establishing the existence of classes or clusters. In contrast, we can know with certainty that there exist some classes, and the aim is to establish a rule able to classify a new register into one of these classes. The first type is known as Supervised Learning and the second one is known as Unsupervised Learning (or Clustering). We believe that the accuracy of the diagnosis in medicine and, in particular, in urology will be improved by using these types of architectures (one supervised and two unsu-

pervised).

With the system of aid to the diagnosis major benefits are obtained both for the patient, by avoiding painful tests, and for the medical centres by avoiding expensive urodynamical tests and reducing waiting lists.

Although the use of ANNs in medicine is a rather recent phenomenon, there are many applications deployed as in the field of diagnosis, imaging, pharmacology, pathology and of course prognosis. ANNs have been used in the diagnosis of appendicitis, back pain, dementia, myocardial infraction (Green, 2006), psychiatric disorders (Peled, 2005)(Politi, 1999), acute pulmonary embolism (Suzuki, 2005), and temporal arteries.

In Urology, prostate cancer serves as a good example of the usability of ANNs (Remzi, 2001)(Batuello, 2001)(Remzi, 2004). However our work is more related with the neurological part which is less explored (Gil, 2006)(Gil, 2005)(Ruiz, 2005).

In this paper we describe the implementations of ANN based systems aiming at support diagnoses of dysfunctions of the LUT. The remaining part of the paper is organized as follows: first, we give a brief description of the employed neural network architectures. Next we describe the design of our proposal and the training of the ANNs by the available data. Then we describe the subsequent testing carried out in order to analyze the results. Finally we draw the relevant conclusions.

2 NEURAL NETWORK ARCHITECTURES

We have tested three different neural network architectures, two unsupervised ANNs and one supervised ANN. The goal is to obtain a system that supports the diagnoses of the dysfunctions of the LUT. The classification in the maps of the unsupervised ANNs and the output from the supervised ANN will assist the urologist in their medical decisions.

2.1 Supervised ANN - Multilayer Perceptron

A typical Multilayer perceptron (MLP) network consists of three or more layers of neurons: an input layer that receives external inputs, one or more hidden layers, and an output layer which generates the classification results (Jiang, 2006)(Fig. 2). Note that unlike other layers, no computation is involved in the input layer. The principle of the network is that when data are presented at the input layer, the network neurons run calculations in the consecutive layers until an output value is obtained at each of the output neurons. This output will indicate the appropriate class for the input data.

The architecture of the ANN (MLP) consists of layer 1, with the inputs that correspond to the input vector, the layer 2, with the hidden layer and the layer 3 which are the outputs (the 3 diagnoses of the LUT) and the learning used is backpropagation and the algorithm runs as follows:

All the weight vectors m are initialized with small random values from a pseudorandom sequence generator. The following steps are repeated until convergence (i.e. when the error E is below a preset value): Update the weight vectors m_i by

$$m(t+1) = m(t) + \Delta m(t) \quad (1)$$

where

$$\Delta m(t) = -h\partial E(t)/\partial m \quad (2)$$

Compute the error E(t+1).
where t is the iteration number, m is the weight vector, and h is the learning rate. The error E can be chosen as the mean square error function between the actual output y_j and the desired output d_j:

$$E = \frac{1}{2}\sum_{j=1}^{n_j}(d_j - y_j)^2 \quad (3)$$

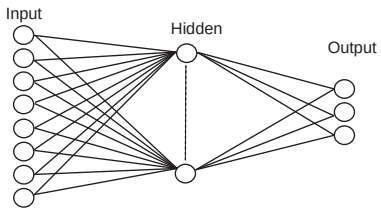


Figure 1: The architecture of the MLP.

2.2 Unsupervised ANN - Kohonen's Self-Organizing Maps

Kohonen's Self-Organizing Map (SOM) is composed of neurons located in a two-dimensional matrix (Kohonen, 1988)(Kohonen, 1990)(Kohonen, 2000). There is a weight vector, $m_i = (m_{i1\ i2}\ ...\ m_{in})$ associated with every neuron in the SOM, where n is the dimension of the input vectors. In our case they are the n fields of each observation of a pattern or a patient in the register.

The SOM is used as a classifier and is organized as indicated in figure 1.

In the fully trained network each neuron is associated with a vector in the input space. The SOM is a soft competitive neural network, which means the winner neuron, i.e. the neuron with the weight vector that is closest to the current input vector according to some measure (dot product in our implementation), gets its weight vector updated so that it becomes more similar to the input vector. The neurons in the vicinity of the winner neuron also get their weight vectors updated but to a lesser extent. Usually a Gaussian function of the distance to the winner neuron is used to modify the updating of the weight vectors. The trained SOM is a projection of the input space, which preserves the topological relationships of the input space. The training of the SOM works as follows:

At time step t an input signal $x \in \mathrm{R}^n$ activates a winner neuron c for which the following is valid:

$$\forall i : x^T(t)m_c(t) \geq x^T(t)m_i(t) \tag{4}$$

The weights are updated according to:

$$m_i(t+1) = \begin{cases} \frac{[m_i(t)+\alpha(t)x(t)]}{||m_i(t)+\alpha(t)x(t)||} & if \quad i \in N_c(t) \\ m_i(t) & if \quad i \notin N_c(t) \end{cases} \tag{5}$$

where $N_c(t)$ is the neighbourhood of the neuron c at time t and $0 < \alpha(t) < \infty$.

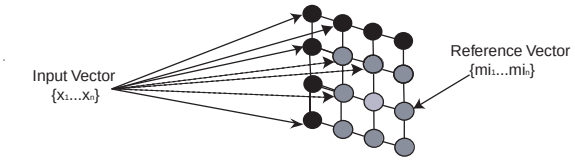


Figure 2: The Kohonen SOM.

2.3 Unsupervised ANN - Growing Cell Structures

It has been pointed out that the predefined structure and size of Kohonen's SOM bring limitations to the resulting mapping. One attempt to solve this problem is the growing cell structures (GCS) (Fritzke, 1993)(Fritzke, 1997), which has a flexible and compact structure, a variable number of elements and a k-dimensional topology where k can be arbitrarily chosen.

In principle, the adaptation of a weight vector in the GCS is done as described in the previous section (SOM), i.e. determine the best matching neuron, adjust its weight vector and the weight vectors of its topological neighbours. However, there are two important differences when compared to the SOM, namely that the adaptation strength is constant over time, and that only the best matching neuron and its direct topological neighbours are adapted.

The GCS estimates the probability density function $P(x)$ of the input space by the aid of local signal counters that keep track of the relative frequency of input signals gathered by each neuron. These estimates are used to indicate proper locations to insert new neurons. The insertion of new neurons by this method will result in a smoothing out of the relative frequency between different neurons. The advantages of this approach is that also the topology of the network will self-organize to fit the input space and the proper number of neurons for the network will be automatically determined, i.e. the algorithm stops when a certain performance criterion is met. Another advantage is that the parameters are constant over time.

The basic building block and also the initial configuration of the GCS is a k-dimensional simplex. Such a simplex is for $k = 1$ a line, for $k = 2$ a triangle, and for $k = 3$ a tetrahedron. In our implementation $k = 2$.

The self-organizing process of the GCS works as follows:

The network is initialized to contain k + 1 neurons with weight vectors $\mathrm{m}_i \in R^n$ randomly chosen. The neurons are connected so that a k-dimensional simplex is formed.

At time step t an input signal $x \in R^n$ activates a winner neuron c for which the following is valid:

$$\forall i : ||x - m_c|| \quad \geq ||x - m_i|| \quad , \tag{6}$$

and the squared distance between the input signal and the winner neuron c is added to a local error variable E_c:

$$\Delta E_c = ||x - m_c||^2. \tag{7}$$

The weight vectors are updated by fractions ε_b and ε_n respectively according to:

$$\Delta m_c = \varepsilon_b(x - m_c) \tag{8}$$

$$\forall i \in N_c : \Delta m_i = \varepsilon_n(x - m_i), \tag{9}$$

where N_c is the set of direct topological neighbours of c. A neuron is inserted if the number of input signals that have been generated so far is an integer multiple of a parameter λ. This is done by dividing the longest edge between the neuron q with the largest accumulated error and its connected neighbour f, and then removing the earlier connection (q, f) and connect r to all neighbours that are common for q and f. The weight vector for r is interpolated from the weight vectors for q and f:

$$m_r = (m_q + m_f)/2. \quad (10)$$

The error variables for all neighbours to r are decreased by a fraction α that depends on the number of neighbours of r:

$$\forall i \in N_r : \Delta E_i = (-\alpha/|N_r|) \cdot E_i, \quad (11)$$

The error variable for r is set to the average of its neighbours:

$$E_r = (1/|N_r|) \cdot \sum_{\iota \in Nr} E_i, \quad (12)$$

and then the error variables of all neurons are decreased:

$$\forall i : \Delta E_i = -\beta E_i \quad (13)$$

3 EXPERIMENTATION

An exhaustive urological exploration with 21 different measurements has been carried out with 200 patients with dysfunctions of the lower urinary tract in order to create a database. The data has been analyzed and processed before entered into the network to ensure that it is homogenized. These 200 registers contribute to the full knowledge adding different values to delimit the ranks of each measure. Each of these registers contains the information measured in the 21 fields showed in Table 1. For this reason, this database plays a crucial role in order to obtain the knowledge base of our system.

The table 1 shows the fields (every variable) of the urological database (their physical units and types of data). This table helps us to understand the dimension of the problem to deal with (different types of data, ranges and incomplete fields). The column direction of the table indicates the meaning for the ANN: all the fields are input for the system except the field diagnosis, which is the dysfunction of each register. There are three dysfunction and one output more for the diagnosis free of dysfunction.

Our database presents a diversity of ranks, values and types. For this reason it is better to start with a process of discretization in order to find a way to guarantee the homogeneity as a first step in the process of training of the ANN. It was adjusted and weighted with the help of the urologists, following their instructions and suggestions.

For example, some of the fields of the database are age, volume of urine and micturition time (as can be seen in table 1). As the differences among all these fields are huge, we created ranks in the values of the fields dividing them into subclasses in order to adjust the data input.

Table 1: The fields of the urological database.

Variable	Type of data	Values	Direction
Age	Numerical	3 85 years	Input
Sex	Categorical	Male, female	Input
Neurological Physical Examination			
Perineal and perianal sensitivity	Categorical	Partial, absence, normal, weak	Input
Voluntary control anal sphincter	Categorical	Partial, absence, normal, weak	Input
Anal tone	Categorical	Normal, relaxed	Input
Free Flowmetry			
Volume of urine	Numerical	7 682 ml	Input
Maximum flow rate	Numerical	4 58 ml/s	Input
Medium flow	Numerical	1 43 ml	Input
Post void residual	Numerical	0 550 ml	Input
Micturition time	Numerical	13 160 s	Input
Cystometry			
Bladder storage	Numerical	50 461 ml	Input
First sensation of bladder filling	Numerical	50 300 ml	Input
Detrusor pressure during filling	Numerical	2 30 cm H20	Input
Test pressure / flow			
Detrusor contraction	Categorical	Invol., Vol., Invol.-Vol.	Input
Volume of urine in micturition	Numerical	0 556 ml	Input
Maximum pressure Detrusor	Numerical	2 200 cm H20	Input
Average flow rate	Numerical	0 10 ml/s	Input
Abdominal pressure	Categorical	Yes, no	Input
Post void residual	Numerical	0 350 ml	Input
Maximum flow rate	Numerical	0 31ml/s	Input
Micturition time	Numerical	2 318 s	Input
Diagnosis			
Areflexia, Hyperreflexia, Obstruction of the LUT, No Dysfunction			Output

For instance:

Age: 0-20 (1), 21-50 (2), 51-65 (3), >65 (4)

Volume of urine: 0-150 (1), 151-300 (2), 301-500 (3), >500 (4)

The numbers between the parentheses are the discretized representations. As we can observe the difficulties are not only in the data ranks, but also in the

types of data, which complicates the process of data discretization.

The next step is to run the experimentation. This has been performed by using cross-validation method. The data has been divided in five sets (S1, S2, S3, S4, S5) and the five experiments performed were:

Experiment 1-Training: T1, T2, T3, T4; Test: V (S5)
Experiment 2-Training: T1, T2, T3, T5; Test: V (S4)
Experiment 3-Training: T1, T2, T4, T5; Test: V (S3)
Experiment 4-Training: T1, T3, T4, T5; Test: V (S2)
Experiment 5-Training: T2, T3, T4, T5; Test: V (S1)

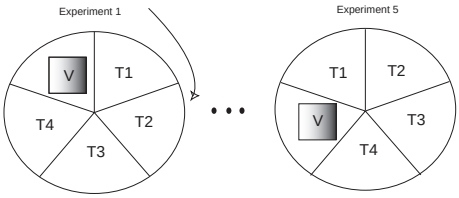


Figure 3: Cross Validation method.

This method is represented in figure 3. There are four data sets used for the process of constructing the model (the training data). The other set of data used to validate the model (the test data). The test data are chosen randomly from the initial data and the remaining data form the training data. The method is called 5-fold cross validation since this process has been performed five times.

The results of the Multilayer Perceptron with the backpropagation algorithm are of around 80% of accuracy. For the other networks, the unsupervised ones, the results are of 76% for the Growing Cell Structures and 74% for Kohonen's Self-Organizing Map. The MLP offers a slightly better performance than GCS and SOM. However the unsupervised ANNs give a useful visual information. This is the reason why we believe a hybrid system with supervised and unsupervised ANNs works properly incorporating the advantages of each ANN.

This comparative and its results with the three different types of ANNs not only has the goal of discovering which is the best neural network but also and mainly to find out relations between the dysfunctions (output of the network). As shown in figure 4 it will help the urologist to pay attention in some tests and its measures to increase the accuracy of the system. In this regard, it will be possible to eliminate some of the tests in order to save money, time and sometimes pain for the patients.

To measure the performance of the ANNs we proceed as follow: Once the ANNs have been trained there are some membership functions for each area, we take a test and we see in which area it is categorized. If the neuron has only input vectors of the same dysfunction the accuracy is 100% otherwise it depends on the mix of dysfunctions and it is another percentage (it incorporates the fuzzy logic idea since it measures the degree of membership to the categorized area). Moreover, a big advantage of using unsupervised ANNs , as we can see in figure 4, is the visual perception of the proximities of the different dysfunctions (their membership functions) categorized in areas to find out relations between similar inputs. The system does not produce negative false. This fact makes it highly reliable amongst the urologists.

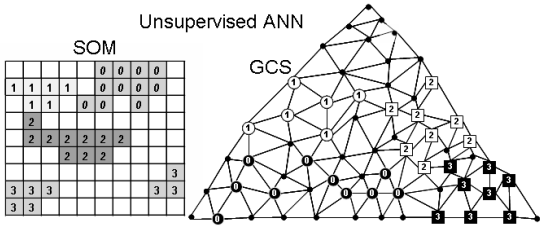


Figure 4: Evaluation of the unsupervised ANNs.

4 CONCLUSIONS AND FUTURE WORK

In this paper we have evaluated the performance of three different kinds of artificial neural networks, the Kohonen Self-Organizing Map, the Growing Cell Structures and the Multilayer Perceptron with the backpropagation algorithm, when applied to the categorization of urological dysfunctions. The ANNs were trained with data from a database with registers of different patients with urological dysfunctions.

The experiment starts with a stage of discretization of the urodynamical measures from a patient in order to provide them to the ANN and to determine if there are any of the three dysfunctions of the LUT or not. In case of finding a dysfunction it would determine what type of dysfunction or dysfunctions the patient could have.

The human expert is able to generalize by using his experience. This big advantage is compensated for ANNs by using graphs offering a visual perception. The frontiers between two dysfunctions help the urologist to discover resemblances. They can also help detecting similarities between fields or urodynamical samples.

In this work we obtained comments valuable from the urologists after using the system. They remarked

its advantages to give a more precise diagnosis and, therefore, to save time and money to the public health. Their comments encourage us to continue our work to develop a system that uses the diagnosis obtained as a result of the combined use of different neural networks. This gives an even more accurate diagnosis.

Next step is the use of data mining involving several steps such as pre-processing with sampling, cleaning and others learning methods as bayesian networks, decision trees, etc.

ACKNOWLEDGEMENTS

We want to express our acknowledgement to Christian Balkenius for helpful comments on the manuscript. The data used in the development of this system is the result of several years of collaboration with urologists of the Hospital of San Juan (Alicante-Spain). The work has been supported by the Office of Science and Technology as part of the research project "Cooperative diagnosis systems for Urological dysfunctions (2005-2006)".

REFERENCES

Abrams, P. Urodynamics. 3rd edn., Springer, 2005,20-39.

Begg, R., Kamruzzaman, J., Sarkar R.: Neural Networks in Healthcare: Potential and Challenges, 2006.

E.H. Shortliffe, Computer-Based Medical Consultations: MYCIN, Elsevier/North Holland, New York, 1976.

Green, M., Bjork, J., Forberg, J., Ekelund, U., Edenbrandt, L. and Ohlsson, M.: Comparison between neural networks and multiple logistic regression to predict acute coronary syndrome in the emergency room. Artificial Intelligence in Medicine, Vol. 38, (2006) 305-318.

Peled, A.: Plasticity imbalance in mental disorders the neuroscience of psychiatry: Implications for diagnosis and research. Medical Hypotheses, Vol. 65, (2005) 947-952.

Politi, E., Balduzzi, C., Bussi, R. and Bellodi, L. Artificial neural networks: a study in clinical psychopharmacology, Psychiatry Research, Vol. 87, (1999) 203-215.

Suzuki, K., Shiraishi, J., Abe, H. MacMahon, H. and Doi, K.: False-positive reduction in computer-aided diagnostic scheme for detecting nodules in chest radiographs by means of massive training artificial neural network. Academic Radiology, Vol. 12, (2005) 191-201.

Remzi M., Djavan B., Seitz C., Hruby S., Marberger M.: Artificial Neural Networks to predict the outcome of repeat prostate biopsies. J Urol (2001).

Batuello JT., Gamito EJ, Crawford ED., Han M., Partin AW., McLeod DG., et al.: Artificial Neural Network Model for the assessment of lymph node spread in patients with clinically localized prostate cancer. Urology, Vol. 57, (2001) 481–485.

Remzi M., Djavan B.: Artificial neural networks in urology. European Urology, (2004) 33-38.

Gil, D., Soriano, A., Ruiz, D., Montejo, C.A.: Development of an artificial neural network for helping to diagnose diseases in urology. Proceedings of the Bio-Inspired mOdels of NEtwork, Information and Computing Systems (BIONETICS'06), (2006).

Gil, D., Soriano, A., Ruiz, D., García, J., Flórez, F.; Decision support system for diagnosing dysfunctions in the lower urinary tract. 3rd European Medical and Biological Engineering Conference. IFMBE European Conference on Biomedical Engineering. (2005).

Ruiz, D., García, J.M., Macía, F., Soriano, A.: Robust Modelling of Biological Neuroregulators. IEEE Engineering in Medicine and Biology Society, (2005).

Jiang, Y., Zheng, J., Peng, C. and Li, Q.: A multilayer perceptron-based medical decision support system for heart disease diagnosis. Elsevier, Expert Systems with Applications, Vol. 30, (2006) 272-281.

Kohonen, T.: Self-Organization and Associative Memory, Berlin Heidelberg, Springer-Verlag, (1988).

Kohonen, T.: The self-organizing map. Proceedings of the IEEE, 78, 9, (1990) 1464-1480.

Kohonen, T.: Self-Organizing Maps, Berlin Heidelberg, Springer-Verlag, (2001).

Fritzke, B.: Growing Cell Structures - A Self-organizing Network for Unsupervised and Supervised Learning, (1993) 93-126.

Fritzke, B.: Unsupervised ontogenic networks. In Fiesler and Beale (eds.). Handbook of Neural Computation, IOP Publishing Ltd. and Oxford University Press, (1997).

SELECTION OF AN ARTIFICIAL NEURAL NETWORK MODEL TO DIAGNOSIS MOUTH-BREATHING CHILDREN

Felipe Mancini, Ivan Torres Pisa
Health Informatics Department, Federal University of São Paulol, Rua Botucatu 862, São Paulo, Brazil
felipe-pg@dis.epm.br, ivan.pisa@unifesp.br

Liu Chiao Yi, Shirley Shizue Nagata Pignatari
Pediatric Otorhinolaryngology Department, Federal University of São Paulol, São Paulo, Brazil
liu@netcomp.com.br, pignatari@terra.com.br

Keywords: Artificial neural networks, posture, mouth breathing, clinical decision support systems.

Abstract: A number of factors can lead to changes in body posture, basically determined by alterations in the natural curvature of the spine. Such changes, in turn, may also result in secondary health problems. Mouth breathing is thought to be one of these problems. Experiments with healthy nasal breathing individuals have showed that when they are forced to breathe through their mouth only the natural shape of their spine curves change. However the characterization of the spine curvature in mouth breathers has not been done yet and the matter lies on the personal experience of the health professional. This study reports on the preliminary findings of a broader research which attempts to characterize the changes in the behaviour of the spine, caused by mouth breathing, by using artificial neural network modelling and data from 52 subjects. Four different models – backprogation, learning vector quantization (LVQ), and self-organizing map (SOM) – were tested for best performances in sensitivity and specificity in diagnosing mouth and nasal breathing children. Competitive-learning-based algorithms – LVQ and SOM – presented the best performance for current data set.

1 INTRODUCTION

Breathing is the first vital function developed at birth becoming the main body function, and as such should be cared for. Chronic mouth breathing is associated with pediatric, allergy-related and otorhinological complaints.

The narrowing of the pharynx has been reported to be associated with forward extension of the neck in the attempt to straighten the pharyngeal tube in order to improve the reduced air flow through it (Solow, 1984).

The skull, mandible, cervical portion of the spine, and upper airways can be viewed as a system in which the positioning of its parts are closely related. Mouth breathing, a physiological change in the correct respiratory process, determines postural changes due to the altered, interrelated performance of the muscles in each of the parts that integrates the above mentioned system (Rocabado, 1979; Ribeiro, 2003). Studies on body posture of mouth-breathing children have reported as characteristic features of these individuals: forward positioned head and shoulders, lordosis, protruded scapulas, frontal depression of the thorax, and protruded abdomen (Aragão, 1991; Liu, 2003). However, no studies further characterizing the posture of mouth breathers are currently available.

A computer-aided tool could be useful for health professionals if it could characterize the postural changes caused by mouth breathing. Artificial neural networks (ANN) (Haykin, 1999) have been used successfully in treating and analyzing biomedical data[7]. ANN can provide a faster data analysis when associations between factors and outcomes are not linear or when there are a great number of factors (high dimensionality) to be analyzed (Lisboa, 2002). Furthermore, ANN can lessen the influence of confounding variables (noise) (Reggia, 1993).

The aim of this study is to report the preliminary findings regarding the selection of an ANN model for identifying mouth-breathing children through the analysis of their posture.

2 MATERIALS AND METHODS

The data used in this study was collected at Imaging Department and Pediatric Otorhinolaryngology Department at Federal University of São Paulo (UNIFESP), Brazil. Fifty two children were assessed including 30 previously diagnosed as mouth-breathing subjects and 22 nasal breathers. The variables collected for analysis are shown in Table 1.

Table 1: Study variables.

Anthropometrics	Diaphragm	Posture
Sex	Right side Excursion (PD)	Cervical Curvature
Age		Lumbar Curvature
Weight	Left side Excursion (PE)	Thoracic Curvature
Height		Pelvis Positioning

The imaging of diaphragm excursion was obtained by videofluoroscopy and recorded for analysis using Adobe Photoshop® (Adobe Systems Inc.) software. Due to the limited size of the fluoroscope screen, the imaging of the left and right sides of the diaphragm was recorded separately. The posture of the participating children was assessed using photographs of the subjects' left-side view on which angles formed by key body features were analyzed using a specially developed software (Software for Posture Evaluation, SAPO) (Duarte, 2006). Figure 1 shows these key features and the angles they determined.

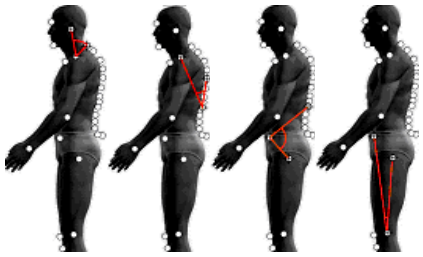


Figure 1: Representation of key points, and respective angles they formed, used in the posture assessment: (a) cervical curvature; (b) thoracic curvature; (c) lumbar curvature; (d) pelvis positioning.

The data collected were used to determine the ANN model that showed the best performance among a number of models which included, backpropagation (BP) (Haykin, 1999), learning vector quantization (LVQ) (Kohonen, 1997), and self-organizing map (SOM) (Kohonen, 1997). Such a comparison was carried out through Matlab® development tool (The MathWorks Inc., Natick, MA, USA) and implementation packages SOM Toolbox package (Vesanto, 2000) for SOM and LVQ models and the Neural Networks Toolbox® (The MathWorks Inc., Natick, MA, USA) for BP model.

The structure of each ANN model was as follows:

- Backpropagation - Network structure: 20 nodes in the first hidden layer, 5 nodes in the second hidden layer and 1 node in output layer. Training function: Levenberg-Marquardt. Maximum number of epochs to train: 100. Minimum performance gradient: 10^{-10};
- LVQ - Network structure: 3 x 3 nodes. Vectors prototypes initialization: linear. Neighbourhood relationship: hexagonal. Running length: 100. Learning rate used in training: 0.001;
- SOM - Network structure: 3 x 3 nodes. Vectors prototypes initialization: linear. Neighbourhood relationship: hexagonal. Neighbourhood function: gaussian.

The performances of these ANN models were measured by determining the rates for sensitivity and specificity obtained with each model, when carrying out the leave-one-out cross-validation (Burnham, 2004). Receiver Operating Characteristic (ROC) (Metz, 1978) curve analysis was used to determine the association ANN model-input pattern with the best performance among those showing higher rates of sensitivity and specificity.

Some investigation of factors potentially associated with the shaping of spine curves – body weight, height and excursion of the diaphragm – was also carried out by inputting them either separately or in combination with the data collected regarding spine curvature (Table 2).

Table 2: Different input patterns used to determine the influence of potential influencing factor on spine curvature. The number between brackets indicates the number of variables of a same subset.

Input Pattern (IP)	IP Label
spine curvatures (4) and diaphragm excursion (2)	IP 1
spine curvatures (4), diaphragm excursion (2), weight (1) and height (1)	IP 2
spine curvatures (4)	IP 3
diaphragm excursion (2)	IP 4
weight (1) and height (1)	IP 5

3 RESULTS

The sensitivity and specificity rates attained with each of the study input patterns and each of the study ANN model varied from 0.76 to 1 and from 0.57 to 0.99 respectively (Table 3).

The areas under the ROC curves for LVQ and SOM inputting IP3 were 0.94 and 0.90, respectively, for SOM inputting IP2 was 0.92 and inputting IP1 0.97.

Table 3: Specificity (sp) and sensibility (se) values calculated through leave-one-out algorithm for different data sets and for all ANN models analyzed in this study.

RNA Models	**SOM**		**LVQ**		**BP**	
Input Pattern	sp	se	sp	se	sp	se
PE1	0.97	0.97	0.99	0.90	0.98	0.94
PE2	0.98	0.95	0.96	0.96	0.86	0.94
PE3	0.98	0.93	0.98	0.97	0.96	0.90
PE4	0.88	0.87	0.88	0.90	0.88	0.90
PE5	0.80	0.76	0.67	0.76	0.67	0.76

4 DISCUSSION

The area under the ROC curve for IP2 (all variables studied) inputted in SOM (0.92) indicates that this input pattern improves the performance of SOM but is still bellow the performance of LVQ using a much simpler set of input variables (IP3).

However, a previous statistical test not shown in this study (t-Student test) presented a statistically significant association between the data collected on the excursion of the diaphragm and the child being a mouth breather. This association seems to reflect on the area under the ROC curve (0.97) calculated for SOM model when the variables associated with spine curvature and diaphragm excursion (IP1) was inputted.

Despite the input of the data referring to the diaphragm excursion (IP1 and IP2) yielding a better performance of SOM, the fluoroscopic investigation is an additional medical examination that is not usually performed in the clinical practice. Therefore, if we are to deal with such limitation, LVQ model associated with the input of variables of spine curvature only (IP3) can presently be a good alternative model due to its high rates of sensitivity and specificity.

Including the variables weight and height to the set spine curvature and diaphragm excursion (IP1) to form IP2 resulted in lower performance of SOM model according to ROC curve analysis. This agrees with previous statistical analysis (Student's t-test) showing that the variation of weight and height between mouth and nasal breathers was not statistically significant.

Pesonen et al. (1996), Markeya et al. (2003), and Ng & Chong (2006) compared the performance of SOM and BP models in different tasks of classification of biomedical data and found that BP had higher rates of specificity and sensitivity. This was not the case in the present study. In fact, training in SOM is unsupervised, which would support its worse performance in data classification as compared with models using supervised training. A potential explanation for the best performance of SOM over BP in the present study is the limited set of data (52 patients) for training and validation currently available.

As previously mentioned, the present report is part of a broader biomedical study. So far, the use of computer-aided modelling focused the development of a reliable diagnosis tool. This is deemed to be the first step to develop a second and perhaps more important tool that could indicate the severity of changes in body posture and assist the decision making regarding the prescription of a physiotherapeutic treatment for such condition. ANN modelling is a resource that could overcome the complexity of such task.

5 CONCLUSIONS

The best rates of sensitivity and specificity were attained for variables associated with the spine curvature only (IP3) inputted in LVQ model. A further comparison of performance using IP3 was carried out between SOM and LVQ models using their respective ROC curves which showed that the area under the curve for LVQ model was larger (0.94) than that for SOM (0.90).

Although supervised learning ANN models, such as BP model, have been reported to yield better rates of sensitivity and specificity, the present study found that SOM and LVQ, both competitive-learning-based algorithms, had better performance.

REFERENCES

Aragão W. *Aragao´s function regulator, the stomatognathic system*. J Clin Ped Dent. 1991; 15(4):226-231.

Burnham, KP, Anderson D. *Model Selection and Multi-Model Inference*. Berlim: Springer; 2004.

Duarte M. *Software for Posture Evaluation (SAPO)*, Brazil: University of São Paulo; 2006.

Haykin S. *Neural Networks: a Comprehensive Foundation.* New Jersey: Prentice-Hall; 1999.

Kohonen T. *Self-organizing Maps*. Berlim: Springer-Verlag; 1997.

Lisboa PJG. *A Review of Evidence of Health Benefit from Artificial Neural Networks in Medical Intervention*. Neural Netw. 2002; 15: 11-39.

Liu CY, Guedes ZCF, Pignatari SSN, Weckx LLM. *Alteração da postura corporal em crianças respiradoras bucais*. Fisioter Mov. 2003;16(3):29-33.

Markeya MK, Lo JY, Tourassib GD, Floyd CE, *Self-organizing map for cluster analysis of a breast cancer database*, Artif Intell Med, 2003, 27, 113–127.

Metz CE. *Basic principles of ROC analysis*. Seminars in Nucl Med. 1978;8:283–98.

Ng EY & Chong C. *ANN-based mapping of febrile subjects in mass thermogram screening: facts and myths*. J Med Eng Technol. 2006; 30(5):330-7.

Pesonen E, Matti E, Juhola M. *Comparasion of Different Neural Network Algorithms in the Diagnosis of Acute Appendicitis*. Int J Biomed Comput. 1996; 40:227-233.

Reggia JA. *Neural Computation in Medicine*. Artif Intell Med 1993; 5: 143-157.

Ribeiro EC, Soares LM. *Avaliação espirométrica de crianças portadoras de respiração bucal antes e após intervenção fisioterapêutica*. Fisioter Bras. 2003;4(3):163-7.

Rocabado M. Cabeza Y. *Cuello – tratamiento articular*. Buenos Aires: Intermédica, 1979.

Solow B, Siersbaek-Nielsen S, Greve E. *Airway adequacy, head posture, and craniofacial morphology*. Am J Orthod. 1984;86:214-23.

Vesanto J, Himberg J, Alhoniemi E, Parhankangas J. *SOM toolbox for Matlab 5*. Espoo: Helsinki University of Technology; 2000.

MIAWARE SOFTWARE
3D Medical Image Analysis with Automated Reporting Engine and Ontology-based Search

Bartłomiej Wilkowski
Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark
bw@imm.dtu.dk

Óscar N. M. Pereira, Paulo Dias
IEETA-Institute of Electronics and Telematics Engineering of Aveiro, University of Aveiro, Aveiro, Portugal
omp@ua.pt, paulo.dias@det.ua.pt

Miguel Castro
Centro Hospitalar do Baixo Alentejo - Hospital José Joaquim Fernandes de Beja, Beja, Portugal
migkastro@yahoo.com

Marcin Janicki
Department of Microelectronics and Computer Science, Technical University of Łódź, al. Politechniki 11, Poland
janicki@dmcs.pl

Keywords: Computed axial tomography, Ontology, Radiological report, Image visualization.

Abstract: This article presents MIAWARE, a software for Medical Image Analysis With Automated Reporting Engine, which was designed and developed for doctor/radiologist assistance. It allows to analyze an image stack from computed axial tomography scan of lungs (thorax) and, at the same time, to mark all pathologies on images and report their characteristics. The reporting process is normalized - radiologists cannot describe pathological changes with their own words, but can only use some terms from a specific vocabulary set provided by the software. Consequently, a normalized radiological report is automatically generated. Furthermore, MIAWARE software is accompanied with an intelligent search engine for medical reports, based on the relations between parts of the lungs. A logical structure of the lungs is introduced to the search algorithm through the specially developed ontology. As a result, a deductive report search was obtained, which may be helpful for doctors while diagnosing patients' cases. Finally, the MIAWARE software can be considered also as a teaching tool for future radiologists and physicians.

1 INTRODUCTION

This article presents MIAWARE, a software, which enables doctors and radiologists to carry out an examination of the patient's lungs condition through the close analysis of the computed axial tomography images and then, in parallel, to perform health state reporting process. Secondly, an intelligent search engine for radiological reports is presented, together with all its advantages over the ordinary searching schemas. The screenshot of the MIAWARE's application graphical user interface is shown in Figure 1. MIAWARE was developed completely in Java programming language together with some embedded native code wrappers used.

Nowadays, it is very common that a radiologist performs the analysis of the radiological images in its own, favorite manner. Some of the radiologists report all pathological changes encountered in the radiological images speaking to a microphone and recording their voice. Afterwards, the recorded tape is listened out and a medical text report is produced. Another radiologists write reports alone in the moment of performing analysis.

There can be found some serious shortcomings in such reporting schemas, which may affect the accuracy of the medical diagnoses. The main problem is that reports differ in structure from radiologist to radiologist. Every human has different way of thinking, different way of expressing things, remarks and

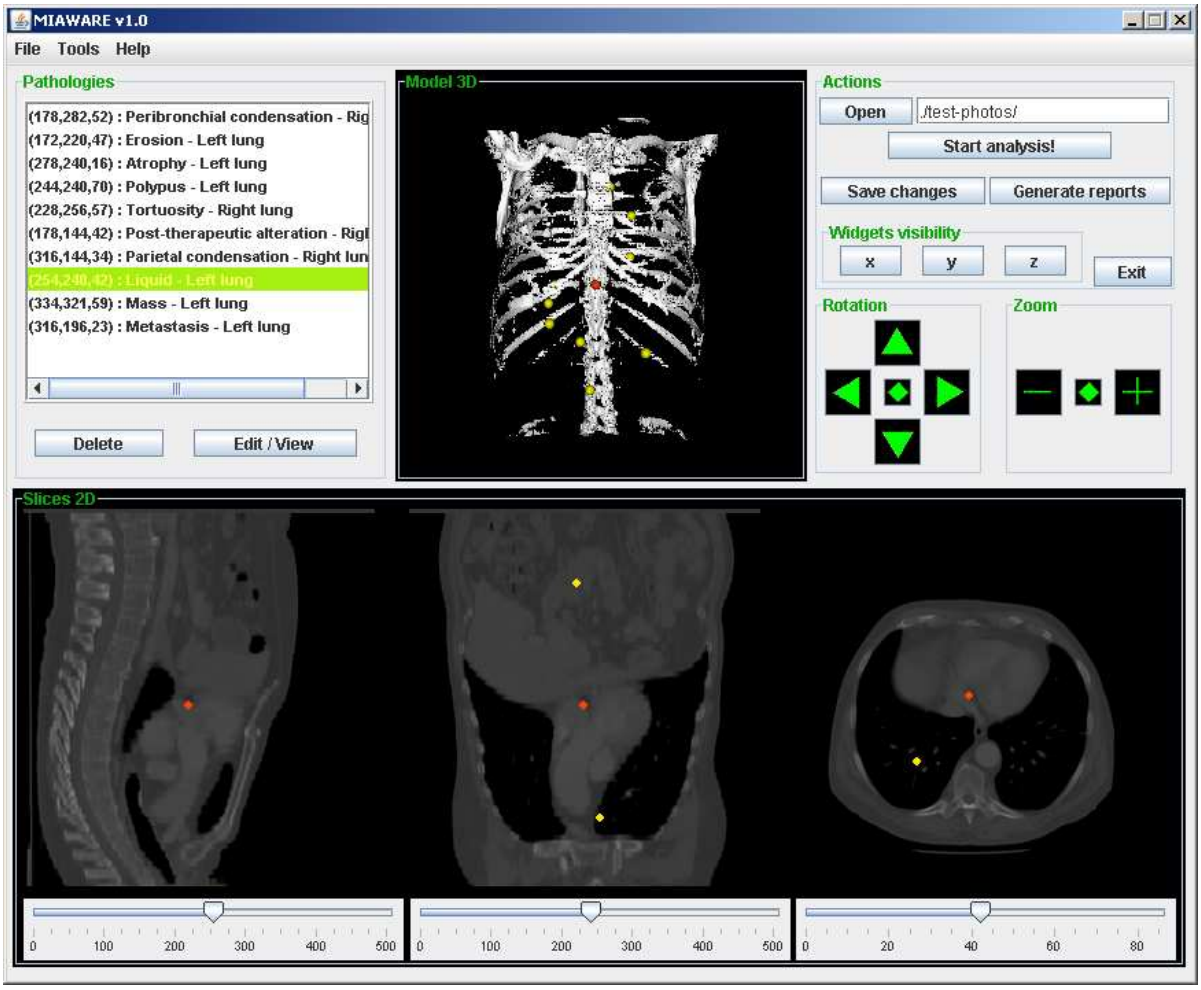


Figure 1: The graphical user interface of the MIAWARE software. The radiologist is able to cut the 3D model in order to find pathologies on the 2D views, mark them by simple mouse click and attach necessary description. All pathologies are listed out in the Pathologies panel. After the examination, the normalized medical report can be generated.

observations. It means that given the same medical data, the same patient's case to many radiologists in order to make analysis, it can and will, almost surely, produce many different reports with different layouts and various observations on the patient's health. As a result, a doctor may interpret each of such reports differently.

This is the reason why MIAWARE software's main objective is to generate medical reports in a normalized way. Such reports should contain only medical data, which describes in details the encountered pathologies using a standardized layout. The radiologist does not use his words in order to report a pathology, but oppositely, he fills up a provided reporting form by choosing suitable medical terms.

Moreover, pathology reporting in MIAWARE is performed in the moment of image analysis using interactive graphical user interface. Radiologist can mark the location of the pathology on the image and associate with this point a necessary description. This allows him to be always focused on the images. Finally, a normalized medical report over all pathological changes is generated by the software. The full description of how a normalized reporting process is performed with MIAWARE software is described in details in section 3.

Normalization of the reports improves significantly its further processing possibilities. One example can be the developed search engine for the MIAWARE medical reports. An efficient search engine for medical reports can be considered as very useful and may help the doctor while making diagnoses.

Further sections will describe in details the architecture of the MIAWARE software and the functionality of its modules.

2 VISUALIZATION OF IMAGES

The entire visualization of CAT scan stack images and 3D model creation is performed using the Visual Toolkit (VTK - www.vtk.org) (Schroeder et al., 1998). VTK is made in C++ language, but it provides suitable wrapper classes for Java. Moreover, ImageJ (http://rsb.info.nih.gov/ij/) software classes are used in order to obtain properties of the analyzed CAT image stack.

MIAWARE graphical user interface provides a 3D view of the radiological stack of images. This is generated using VTK wrapper classes, which create a special pipeline. After loading image data into memory, a contour filter is applied followed by proper mapping of polygonal data and graphics primitives. Finally, a 3D actor is added to the special panel, which actually is a rendering window for three-dimensional scene.

The visualization in MIAWARE consists also of three 2D cross-sectional image views. They are generated by three widgets (interactive window objects), present on the 3D scene, which are able to cut the model in three plane directions and provide 2D image data for the cross-sectional views (see Figure 2). Widgets can be easily moved by the radiologist along its respective direction axis in order to perform model cutting. It should be mentioned, that CAT scan provides the radiologist with image stack in axial plane. The image data for two other 2D planes and the 3D model are obtained and rendered by the software after a proper initial stack data processing.

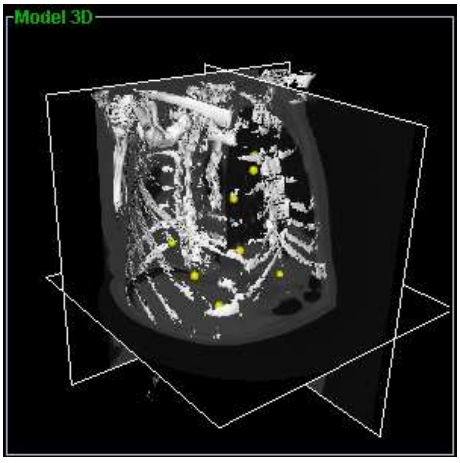


Figure 2: A view of the 3D model with three visible cutting widgets.

Finally, the panels, which display cross-sectional views of the model are enhanced with a very important feature. Radiologist is able to mark any pathology, encountered during the analysis, directly on the 2D view by a simple left mouse button click over that location. The clicked point is automatically marked on all three cross-sectional views (as a yellow circle) and 3D scene (yellow sphere). Afterwards, the radiologist is able to attach precise information and description of that physiological change to the marked point. The description of how the pathology information is defined and added to the specified location is presented in section 3. It should be also remembered that all the pathologies defined by the radiologist can be saved and retrieved for further analysis of the same CAT stack.

3 PATHOLOGY REPORTING

As it was mentioned in the introductory part of this article, the reports generated with MIAWARE software are normalized. This is achieved thanks to a special pathology reporting form implemented in this software. According to the previous section, radiologist is able to mark any location on the 2D image in order to define and describe the encountered pathology. Such an information is added through a combobox-based form, which provides radiologist with medical terms necessary for an effective name, type and pathology location specification. The most innovative here is the fact that the radiologist cannot describe those findings with his own words, but can use only the specific medical vocabulary provided by the application. Consequently, MIAWARE software is able to create normalized medical reports according to the information about all pathologies introduced earlier by the radiologist.

Arrangement and selection of the vocabulary was made after the consultation with doctor Miguel Castro working in hospital in Beja (Centro Hospitalar do Baixo Alentejo - Hospital José Joaquim Fernandes de Beja) and RadLex (A Lexicon for Uniform Indexing and Retrieval of Radiology Information Resources) term browser, which can be found on the Radiological Society of North America web page (RSNA.org, 2007). RadLex term browser was created in order to unify the radiological vocabulary used during image analysis and reporting procedures.

The entire vocabulary is kept in the XML file together with a declaration of the vocabulary for all comboboxes (set of medical terms), which are presented to the radiologist during the pathology definition. A vocabulary set presented in any subsequent combobox is dependent of a previous radiologist's choice. For example, if radiologist has defined that the pathology is located in the left lung, the next combobox will offer him to choose all subparts (lobes) of left lung. The example pathology definition steps in

Table 1: Example pathology definition steps in MIAWARE.

Step	Combobox title	Selected value
1.	Morphophysiological process	Neoplastic process
2.	Neoplastic process	Mass
3.	Location	Left lung
4.	Left lung location	Upper lobe
5.	Left lung upper lobe location	Lingula
6.	Left lung upper lobe lingula location	Superior segment

MIAWARE is presented in Table 1.

When the analysis of the CAT stack is finished, radiologist is able to generate a final medical report over all the pathologies already defined. It is done by pressing the *Generate reports* button. This action produces reports in two formats: Plain Text Format (TXT) and Rich Document Format (RDF - www.w3c.org/RDF/), computer understandable format. The first one can be verified and analyzed later by the doctor in order to make diagnosis. The generated text report has a well defined structure and its layout differs significantly from the recently created reports. The format of medical reports requires still some discussion over its layout and the ways how it should be created. MIAWARE text report format is only a suggestion, which is intended for further improvement and development. A short fragment of the sample MIAWARE text report is presented here:

```
**** MIAWARE REPORT *******
Generation date: Jun/27/2007

*****************************
Control Point no. 1 : (x,y,z) = (178,282,52)
Specifications:
Morphophysiological process: General process
General process: Peribronchial condensation
Location: Right lung
Right lung location: Lower lobe
Right lung lower lobe location:
->Lateral basal segment

*****************************
Control Point no. 2 : (x,y,z) = (172,220,47)
Specifications:
Morphophysiological process: General process
General process: Post-therapeutic alteration
Location: Right lung
Right lung location: Middle lobe
Right lung middle lobe location:
->Medial segment
...
****** END OF REPORT ******
```

The second type of reports, in RDF format, is created for further processing of its content (report searching). It is described in details in section 4.

4 MEDICAL REPORT SEARCHING

This section describes the structure of the RDF medical reports and the MIAWARE search engine together with an ontology for lungs developed specially for this purpose.

4.1 RDF Reports

As it was already mentioned, the RDF format for medical reports is required for further information processing and searching. RDF model introduces description of resources by statements and its data model contains of three components: resources, properties and statements (called as triples). Resources are any datatype items, which can have any value definition (statement) through some given relation (property). Any statement can consist of a new triple resource-property-statement. "Just as an English sentence usually comprises a subject, a verb and objects, RDF statements consist of subjects, properties and objects" (Gomez-Perez et al., 2004).

Table 1 represents one example of pathology definition. The final medical report will usually contain more such definitions grouped in some specific way. The data gathered in the Table 1 can be represented as normal, lexical group of sentences describing any pathology found. For example:

> 'A morphophysiological process was found. It is in the form of a neoplastic process of the type mass. It is located in the left lung, in its upper lobe, exactly in the superior segment of the lingula.'

Such a group of sentences can be represented as resource-property-statement model and is used in MIAWARE medical RDF reports. In this case, the first underlined word is a resource and the rest is a statement. As our statement consists of group of resources it has to be analyzed further. Then the first resource of the previous statement is a resource and

the rest group another statement. Such embedded structure of the resource-property-statement is created through RDF reified statements. It should be mentioned that the properties (which connect resources with the statements) in the above example are: *in form of*, *of the type*, etc.

The 3RDF reports generated by MIAWARE software keep the pathology information in the manner presented above. It should be only mentioned that the role of properties in our reports play titles of the subsequent comboboxes. These names are taken from the XML file used by pathology definition form, described in section 3.

4.2 Ontology-based Report Searching

The radiological examinations are carried out quite often in such places as hospitals, private and public surgeries or any other medical institutions. As a result, it produces a great amount of medical reports in relatively short time. Such reports should be kept and gathered together for future usage as references to previously encountered and defined pathologies or diseases. Manual searching of great amount of documents is time consuming. As a result, an intelligent search engine of medical reports can significantly speed-up the disease recognition process, as, considering given criteria, it would immediately result in sets of references to the archive reports with similar pathological symptoms in other patients, the resultant diagnoses and applied treatments.

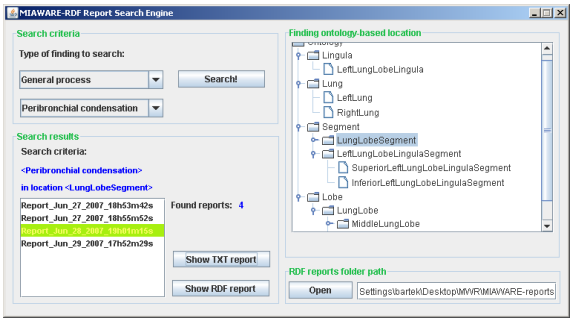


Figure 3: The graphical user interface of the MIAWARE intelligent search engine. The sets of RDF medical reports generated by MIAWARE software can be filtered out according to the given criteria: position in lungs and pathology type. The searching process is performed on a logical basis (not in a lexical way, as ordinary search engines).

The search engine for medical reports developed together with MIAWARE software (Figure 3) is able to find all reports where exist some specified pathology defined in a lung part (specified as a search criteria) or any of its subparts. This adds some intelligence to the searching process, what is explained on the simple example. Let's suppose that a doctor wants to find all reports with a definition of a tumour (first search criterion), which had been found in left lung. Let's have a report with two pathologies defined:

- Polypus in Right lung
- Tumour in Lung lingula

An ordinary lexical search will respond that this report does not match search criteria as the first pathology is not a tumour and the second pathology, which is a tumour, is not located in left lung. Oppositely, the MIAWARE search engine will accept this report as matching the given criteria, because it can deduce that Lung lingula is a subpart of the left lung. Such a logical deduction is performed by our search engine thanks to the lungs ontology, which defines and provides the part-whole relations between the elements of the lungs. This ontology was developed using Jena (http://jena.sourceforge.net/) and Protégé (http://protege.stanford.edu/) software. Sample visualization of the taxonomy of the classes taken from our lungs ontology is presented in Figure 4.

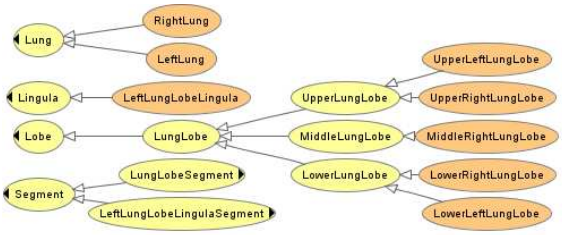


Figure 4: Lungs ontology - hierarchy of classes (created with OWLViz (Drummond, 2007)). Such a taxonomy of classes, related to each other through the specific properties defined in an ontology, allow the search engine to make intelligent (logical) decisions.

Our ontology was developed based on the following article references: (Mejino Jr et al., 2003) (Donnelly et al., 2005) (Guarino and Welty, 2000) (Guarino and Welty, 2004) (Guarino et al., 2000) (Guarino and Welty, 2002) (Michael et al., 2001) (Knublauch, 2004) and book positions: (Gomez-Perez et al., 2004) (Horridge, 2004). Moreover, the information about anatomical structure of the lungs was taken from *Anatomy and Physiology* book (Seeley et al., 2005). The ontology is made in Web Ontology Webpage - OWL (www.w3c.org/2004/OWL/) format.

The search algorithm takes as the criteria the name of the pathology and its location in the lungs. Next, it deduces from the ontology all the subparts of the given lung location and performs comparison of every single pathology description taken from any medical report (in RDF format) with the search criteria. If

there is at least one such a pathology definition which agrees with criteria, it displays respective report as a result. Consequently, the doctor can view and read such a report very easily. We suppose that such filtering of radiological reports may improve doctor's diagnosis and speed-up his decisions.

5 CONCLUSIONS

The presented software is only a first prototype and needs many improvements to be useful in a real context. One of the reason for this is the fact that the vocabulary used during pathology reporting is not sufficient and requires significant expansion and redefinition. However, this software can be considered as a strong fundament for future development in order to achieve a fully operational version.

The ideas presented herein are considered as a potential improvement for image-based medicine and radiological analysis course. MIAWARE software facilitates radiologists with simultaneous analysis of the CAT stack images and pathology reporting without looking away from the monitor. Consequently, the radiologist can be concentrated all the time on the examined images. Moreover, pathologies can be marked on the images and possess the necessary characteristics of respective pathology.

Furthermore, the radiological reports generated with MIAWARE software are always normalized, keeping identical structure and layout independently on the person who performs the analysis. Such a normalization, may help the doctors in better understanding of the reports and it makes room for further report processing and searching.

The intelligent search engine allows rapid medical reports filtering according to the pathologies defined in there. Providing MIAWARE search engine with the knowledge about the parts relationship in the lungs, it is able to deduce internal elements of the specified lung part and to perform report searching of the pathologies not only in the determined lung location, but also in its subparts. This can actually be described as a logical searching of pathologies in the medical reports.

All the features presented by MIAWARE software can lead to the assumption that their implementation into real life may result in more efficient medical diagnosis and faster disease recognition process. Moreover, MIAWARE can be used for investigation and teaching of normalized reporting processes, pathologies and findings classification, statistical processing, etc. Thanks to that, the future radiologist could get their degree through intensive practice with real cases. Finally, the reports generated by the students using MIAWARE software could be evaluated in an automatic manner.

ACKNOWLEDGEMENTS

This work is supported by Lundbeckfonden through the program www.cimbi.org.

REFERENCES

Donnelly, M., Bittner, T., and Rosse, C. (2005). A formal theory for spatial representation and reasoning in biomedical ontologies. *Artificial Intelligence in Medicine*, 36:1–27.

Drummond, N. (2007). Owlviz - ontology visualization tool webpage.

Gomez-Perez, A., Corcho, O., and Fernandez-Lopez, M. (2004). *Ontological Engineering : with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. First Edition (Advanced Information and Knowledge Processing)*. Springer.

Guarino, N. and Welty, C. A. (2000). A formal ontology of properties. *Lecture Notes in Computer Science*, 1937:97–112.

Guarino, N. and Welty, C. A. (2002). Evaluating ontological decisions with ONTOCLEAN. *Communications of the ACM*, 45:61–65.

Guarino, N. and Welty, C. A. (2004). An overview of ontoclean. *Handbook on ontologies*, pages 151–172.

Guarino, N., Welty, C. A., and Partridge, C. (2000). Towards a methodology for ontology based model engineering. *IWME-2000 - International Workshop on Model Engineering*.

Horridge, M. (2004). *A Practical Guide To Building OWL Ontologies With The Protege-OWL Plugin*. University of Manchester, 1 edition.

Knublauch, O. H. (2004). Weaving the biomedical semantic web with the protege owl plugin.

Mejino Jr, J. L., Agoncillo, A. V., Rickard, K. L., and Rosse, C. (2003). Representing complexity in part-whole relationships within the foundational model of anatomy. *American Medical Informatics Association Annual Symposium*, pages 450–454.

Michael, J., Mejino Jr, J. L., and Rosse, C. (2001). The role of definitions in biomedical concept representation. *American Medical Informatics Association Fall Symposium*, pages 463–467.

RSNA.org (2007). Rsna - radlex term browser webpage.

Schroeder, W., Martin, K., and Lorensen, B. (1998). *The Visualization Toolkit- An Object Oriented Approach to 3D Graphics. 2nd ed.* Prentice Hall.

Seeley, R. R., Stephens, T. D., and Tate, P. (2005). *Anatomy and Physiology*. McGraw-Hill Higher Education.

AUTOMATED QUESTION-ANSWERING TECHNIQUES AND THE MEDICAL DOMAIN

Andrea Andrenucci
Department of Computer and System Sciences, Stockholm University/ Royal Institute of Technology
Forum 100, SE-16440, Kista, Sweden
andrea@dsv.su.se

Keywords: Automated Question Answering, Natural Language Interfaces, Medical Applications.

Abstract: The question-answering (QA) paradigm, i.e. the process of retrieving precise answers to natural language (NL) questions, was introduced in late 1960-ies and early 1970-ies within the framework of Artificial Intelligence. The advent of WWW and the need to provide advanced, user-friendly search tools has extended the QA paradigm to a larger audience of people and a larger number of fields, including medicine. This paper reviews and compares three main question-answering approaches based on Natural Language Processing, Information Retrieval, and question templates, eliciting their differences and the context of application that best suits each of them within the medical domain.

1 INTRODUCTION

The question-answering (QA) paradigm, i.e. the process of retrieving precise answers to natural language (NL) questions, was introduced in late 1960-ies and early 1970-ies within the framework of Artificial Intelligence. From the beginning it was mainly an academic research field and there were hardly any commercially applicable QA applications. The advent of WWW and the need to provide advanced, user-friendly search tools has extended the QA paradigm to a larger audience of people and a larger number of fields, including medicine, since medical content is one of the most retrieved types of information on the WWW.

This paper discusses which of three major QA approaches, i.e. deep Natural Language Processing (NLP), Information Retrieval (IR) enhanced by shallow NLP, and Template-based QA, better fit medical applications, eliciting their context of pertinence. To our knowledge, this is the first formal comparison of the three QA approaches that focuses on the medical domain.

The next three sections discuss the approaches and provide some examples of their application in the medical domain; section five and six pinpoints the application areas that fit each technique.

2 NATURAL LANGUAGE PROCESSING (NLP)

A common feature of deep NLP systems is that they convert text input into formal representation of meaning such as logic (first order predicate calculus), semantic networks, conceptual dependency diagrams, or frame-based representations (Jurafsky and Martin, 2000, p. 502). In other words deep NLP systems perform a semantic analysis of text in NL. Semantic analysis is the process of studying the meaning of a linguistic input and giving a formal representation of it.

Jurafsky and Martin (2000, p. 548) provide a possible approach for semantic analysis (see figure 1 on the next page): the user input is first passed through a syntactic parser, whose output, represented with a parse tree, is then processed by a semantic analyzer which delivers a meaning representation.

A medical QA system that implements this approach is the ExtrAns system (Rinaldi et al., 2004). The system derives logical representations of both user questions and the documents in the collection. The documents are analysed in an off-line stage and their semantic form is stored in a Database. In an on-line stage user questions are converted into their semantic representation, prior to being compared to the representations of the documents in the matching process. When a match

occurs, the sentences that originated the match are extracted as possible answers to the user question.

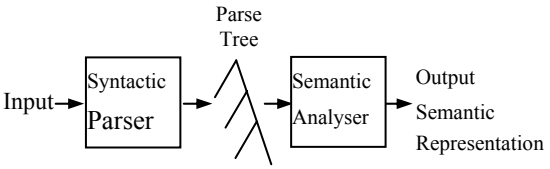


Figure 1: The steps in Semantic Analysis.

Drawbacks of the deep NLP approach are its computational intensiveness and its high processing time (Andrenucci and Sneiders, 2005, Rinaldi et al., 2004) as well portability difficulties (Andrenucci and Sneiders, 2005, Hartrumpf 2006). Figure 2 (Androutsopoulos, Ritchie, and Thanisch, 1995) shows the possible architecture of a typical deep NLP QA system. Six components (linguistic front-end) change when the input language changes, and three components (domain-dependent knowledge) change when the knowledge domain changes. The domain dependent knowledge contains information specific for the domain of interest: a lexicon and a world model. The lexicon contains admissible vocabulary words from the knowledge domain. The world model describes the structure of the domain of interest, i.e. the hierarchy of classes of the domain objects, plus the properties and the constraints that characterize the relationship between them. The linguistic front-end parses and analyses the user input in NL.

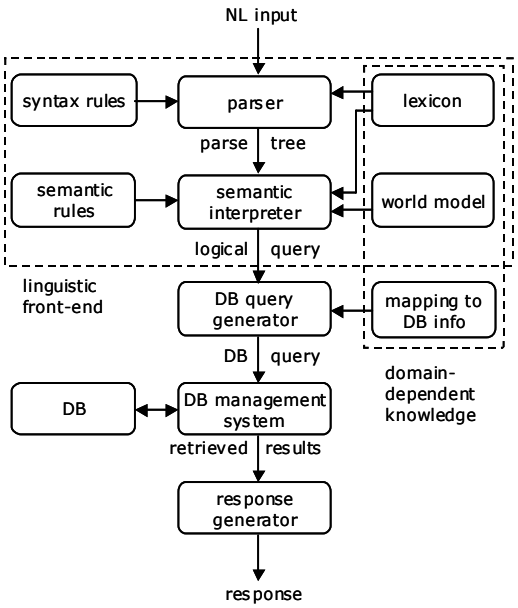


Figure 2: Architecture of a typical deep NLP system, originally from Androutsopoulos et al., 1995.

3 INFORMATION RETRIEVAL (IR) AND SHALLOW NLP

IR has evolved from document retrieval systems to passage retrieval systems, which focus on retrieving text passages rather than entire documents. Answers are extracted with the help of *shallow NLP*, which does not imply text understanding, i.e. semantic analysis of NL input. Instead it focuses on extracting text chunks, matching patterns or entities that contain the answer to user questions. For instance in a question like "Who discovered the polio vaccine?", the presence of the interrogative pronoun "who" implies the extraction of an entity of type "person name" associated with the keywords, "discovered", "polio", "vaccine".

This approach has been implemented in several biomedical systems. Rindflesch et al. (2000) utilized named entity recognition techniques to identify drugs and genes in biomedical documents, then the keywords which connect them (predicates). Craven and Kumlien (1999) utilized "bag of words" at the sentence level, to extract relations between proteins and drugs from the information stored in Medline articles (MEDLINE, 2006).

The IR approach is more domain-independent than traditional NLP, but requires the answer to be explicitly present in the text (Voorhees, 2001). Furthermore answers retrieved with IR techniques are less justified by the context, since they only focus on extracting text snippets containing words that are present in the user question (Laurent, Seguela and Negre, 2006).

This approach is typical for information extraction and is largely used in the Text REtrieval Conferences (Voorhees, 2001), which aim at comparing QA systems that retrieve mainly factoid questions. Several systems that implement the shallow NLP approach exploit data redundancy (Brill et al., 2001), i.e. a number of text passages that contain similar statements, in order to find reliable answers. For example Sekimizu, Park and Tsujii (1998) exploits domain specific verbs, which occur frequently in MEDLINE abstracts, in order to locate the biomedical terms that are respectively subject and object terms for the verbs, and thereafter classify their relations (e.g. Protein X *regulates* Protein Y). Similarly Spasic, Nenadic, and Ananiadou (2003) measure frequent co-occurrences of biomedical verbs with unclassified terms in order to extract domain specific terms.

A medical search system that implements both IR techniques and deeper NLP techniques is PERSIVAL (McKeown et al., 2001). The system

supports user search and summarization of medical information with the help of representations of medical texts and patient records. The system processes medical documents with part of speech tools and with a finite state grammar (that regulates syntactic constraints) in order to extract multi-word terms. This step is similar to the syntactic analysis provided with the help of syntax rules in fig. 2.

Also similarly to the deep NLP approach, this system utilizes a well defined world model (see section 2), provided by the UMLS medical knowledge base (McCray and Nelson, 1995), in order to define the semantic category and the level of specificity of the extracted terms. This is a kind of semantic analysis.

The user profiles and the medical documents are represented with vectors, which are typical IR representation models. The vectors enclose the semantic categories of the medical terms and their associated values. The representations are then compared calculating the cosine similarity of the vectors (Salton and Buckley, 1988), which is also a typical IR technique. Tests conducted with the system (Teufel et al., 2001) have shown that the semantic analysis enhances precision and recall of the system, compared to standard IR techniques.

4 TEMPLATE-BASED APPROACH

Template-based QA extends the pattern matching approach and exploits a collection of manually created question templates, i.e. questions which have open concepts to be filled with data instances, mapped into the conceptual model of the knowledge domain. The templates generally cover the most frequently asked questions (FAQs) of the domain (Sneiders, 2002b), and can be either static, where each template is a question linked to a piece of static text, or dynamic and parameterized if they cover a structured database (Sneiders, 2002a). A question template is viewed as a predicate with variable and fixed parameters:

$\exists\ data_1,\ldots data_n$:Q(fixed$_1$,…fixed$_m$,variable$_1$,… variable$_n$)

During the process of matching a template to a user question, the fixed parameters (fixed$_1$,…fixed$_m$) are bound to the user question. If there are database data instances ($data_1$,…$data_n$) that fit the variable parameters (variable$_1$, …variable$_n$) and make the statement Q true, then these data instances constitute the answer. This approach has been utilized on a medical portal aiming at providing cross language QA in matters of psychology and psychotherapy (http://www.web4health.info).

A similar approach, which focuses on classifying user questions with the help of pre-determined semantic patterns, is applied in a feasibility study for creating a QA prototype for the oral surgery domain (Jacquemart and Zweigenbaum, 2003). The patterns are created with triples that contain two concepts and their relation (Concept A – Relation – Concept B). The relation between the concepts is defined with the help of the UMLS Semantic Network (McCray and Nelson, 1995).

The Medline Button system (Cimino et al., 1992) tries to automate the question generation process creating semantic patterns of concepts that occur frequently in user questions. The system then instantiates the generic concepts in the templates with terms that are specific for the search context and user interests. For instance the template "Does <*procedure*> cause <*disease*>?" is instantiated to "Does *chest x-rays* causes *cancer*?" if the user is interested in those topics.

The PICO-format (Sackett et al., 2004), utilized in several medical QA systems (Niu et al., 2003, Demner-Fushman and Lin, 2005), consists of templates that classifies NL input with the help of a conceptual structure that represent the key elements of clinical questions: Problem (the primary problem of the patient), Intervention (medication or therapeutic procedure), Comparison (of the actual intervention to other possible interventions) and Outcome (the effect of the intervention).

A system that implements IR and templates-based techniques is the EPoCare QA system (Niu et al., 2003). Candidate answers are first retrieved with standard IR techniques and then classified with the PICO format, prior to being matched to PICO-formatted user questions. The system also tries to classify the relations between the PICO conceptual units, for instance individuating cause-effect relations between interventions and outcomes.

5 QA TECHNIQUES AND THE MEDICAL DOMAIN - DISCUSSION

As mentioned in section 3, the IR approach distinguish the expected answer type (e.g. person, place or time) with the help of the so-called "wh-words" in the user question (e.g. who, where, when). Niu et al. (2003) states that this classification is not

appropriate for the medical domain for the following reasons:

1) Questions about patient care usually deal with diagnosis, treatments, prognosis and outcome of the treatments (Richardson et al., 1995). This require a methodology for identifying answer types that is different from the traditional approach utilized for generic "factoid" QA systems.

2) Answers to "when"-questions in medical area are usually related to relative time (Q: "When should I eat my medicine?" A: "One hour before lunch") rather than absolute time/dates, which is typical for generic QA systems (Q: "When was America discovered?" A: "1492"), or address a clinical condition (Q: "When should I see a therapist?" A: "You should consider professional advice if your personal problems are affecting your quality of life and social functions at work or at home for more than a month"). This requires a deeper semantic interpretation of the user question.

3) Yes-no questions, i.e. questions that require yes or no answer (e.g. "Is cognitive behaviour a good therapy method for a person suffering from anxiety disorder?") are not considered by systems that focus on "wh"-questions.

Furthermore IR techniques extract answers containing words that are present in user questions, without considering contextual information in the text that could be relevant to provide and justify answers (Niu and Hirst, 2004, Laurent, Seguela and Negre, 2006). In medical applications correct answers may be missed or incorrect answers may be retrieved if contextual information is not understood, since the context may provide more evidence, clarify or even contradict the extracted snippets (Niu et al., 2003).

Deep NLP-based and Template based QA are the techniques that better fit QA in medical matters. Both approaches handle more advanced types of questions that implies understanding of their context, such as yes-no questions, and have shown better results when it comes to requests for "advice-giving" (e.g. "How to…" questions) since they perform a semantic interpretation of user input (Andrenucci and Sneiders, 2005, Laurent, Seguela and Negre, 2006). For example semantic analysis proved to improve the disambiguation of causal questions that involves reasoning (e.g. "How does cancer develop?") and the precision results of question answering (Girju, 2003).

In the template based approach the interpretation is done manually, individuating for each single template the concepts that cover a part of the conceptual model of the knowledge domain. In the NLP approach the interpretation is done automatically by the system as questions are asked, mapping user questions and candidate answers into a formal semantic representation.

Unlike IR enhanced by shallow NLP, those techniques do not rely on data redundancy, which is more likely to be useful in large, open domains (Molla et al., 2003). Information in restricted domains, such as the medical one, is usually well structured and it is unlikely that answers to the same question are redundantly present in several places of the information source (Niu et al., 2003). Deep NLP and Template based QA are the techniques that are more often utilized to form interfaces to structured data (Andrenucci and Sneiders, 2005).

However there are some important differences that determine the context of application of the two afore-mentioned techniques. The NLP approach provides a natural flow in the user-computer dialogue that resembles human-to-human communication, thanks to the implementation of realistic discourse planning models; see for instance (Buchanan et al., 1995). NLP-based systems may also implement dialectical argumentation techniques in order to be more persuasive while giving advice in health matters. One example is the DAPHNE system (Cawsey, Grasso, and Jones, 1999), which provides advice for the promotion of healthy nutrition and implements a persuasive conversational model based on providing supports for its claims ("People who eat more fruit have less diseases") and anticipating possible counter arguments and exceptions ("Although you may not like all types of vegetables…").

So in dialoguing or counselling matters that have to resemble the patient-doctor communication, the NLP approach is preferable. The NLP-approach also delivers more reliable answers in comparison to the other approaches (Andrenucci and Sneiders, 2005, Molla et al., 2003, Teufel et al., 2001). For example in Power Answer (Moldovan et al, 2003), the best performing system for TREC 2004 and 2005, a logic proof based on abductive justifications is performed among the candidate answers prior to presenting the valid answers to the users, enhancing the quality and reliability of the results. Power Answer achieved an accuracy of 70% while other medical systems implementing approaches similar to the template based approach achieved 60% (Jacquemart and Zweigenbaum, 2003) of accuracy. Among IR systems, Persival (McKeown et al., 2001) achieved precision results that varied between 65 % and 89 %, however IR techniques were supported by syntactic and semantic analysis.

So in cases where the reliability of the answers is vital, systems enhanced by NLP approach are preferable; for instance medical systems that support practitioners in their decision making process and that provide evidence for the suggested answers (Lin and Demner-Fushman, 2005); the so-called evidence-based medicine (Sackett et al., 2000).

A major drawback of this approach is that development and maintenance of NLP systems are complex and require highly qualified personnel such as programmers, knowledge engineers and database administrators. For example when the NLP QA needs to be adapted to multi-lingual environment, changes needs to be applied to the whole linguistic module, which includes the lexicon, the world model, the semantic interpreter and the syntactic parser (Androutsopoulos, Ritchie, and Thanisch, 1995). Another drawback is that this approach is computationally intensive and requires high processing time, which makes it difficult to adapt it to the Web (Rinaldi et al., 2004, Hartrumpf, 2006).

Template-based question answering is the most viable approach when it comes to medical information portals on the Web (Andrenucci and Sneiders, 2005). This is due to the following characteristics: 1) its suitability to support multi-lingual content, 2) the relatively easiness of maintenance, 3) its capacity to solve linguistic ambiguities such as word sense disambiguation without computationally expensive software, 4) and its capability to return answers in different formats.

The suitability to support content in several languages has a simple explanation: user questions are matched against question templates that match different interpretations of the same question and contain individual lexicons; this implies that it is only necessary to change individual templates to get a multi-lingual matching.

Template-based QA systems are also easier to manage since they do not require rare skills: the administrator must only have knowledge of the subject domain and possess basic linguistic skills (Sneiders, 2002a).

Thanks to the usage of multiple lexicons, i.e. small individual lexicons attached to each template, linguistic ambiguities are solved at the micro-level rather than at the macro-level. Small lexicons identify mutually exchangeable words (synonyms and their grammatical forms) for every concept within the narrow context of a given template/ document, rather than in the context of the whole knowledge domain (Sneiders, 2002a), which is typical for the deep NLP approach. This makes the individuation of word meanings in different contexts an easier and less error-prone process (Sneiders, 2002b, p. 262).

The template-based approach supports also the retrieval of answers in a variety of multimedia forms, such as spoken languages, audio-files and imagery (Andrenucci and Sneiders, 2005, McKeown et al. 2001).

The Template based approach has a high recall level, which fits users who are interested in retrieving complete sets of answers rather than few very precise answers.

A drawback of this approach is that manual creation of the templates is required. This is a tedious process, which poses great consistency demands among the persons who create the templates. Another drawback is that the template-based QA does not provide a natural flow in user/system dialogue or provides dialogues of poor quality. One of the first medical systems trying to use templates while dialoguing with users was Eliza (Weizenbaum, 1966), a conversational agent created to simulate the responses of a psychotherapist. The system did not contain any domain knowledge and the templates utilized regular expression in order to match user input and to create responses that exploited keywords from the input sentences. This resulted often in nonsense answers and nonsense dialogues (Copeland, 1993).

6 CONCLUSIONS

This paper has discussed three main techniques within QA and has pointed out the approaches that are more suitable for medical applications: the deep NLP approach and the template based approach.

The template based approach is the most viable commercially and fits Web-based medical applications that are aimed at retrieving multilingual content in different multimedia formats. Its high recall level makes it the technique that fits users who are more interested in retrieving complete sets of answers rather than few very precise answers.

The deep NLP approach provides a dialogue that better resembles the human-to-human conversation and also delivers more reliable answers. It fits areas where the precision of the retrieved information is crucial, e.g. in decision-support or evidence-based medicine.

IR enhanced by shallow NLP is more appropriate as a search tool for larger or open domains as the Web, since it exploits data redundancy. However it can mainly retrieve factual answers unlike the NLP and the template based approaches, which support more complex types of questions such as requests for “advice-giving”.

REFERENCES

Andrenucci, A. and Sneiders, E., 2005. Automated Question Answering: Review of the Main Approaches. In *ICITA'05*, IEEE press.

Androutsopoulos, I., Ritchie, G. and Thanisch P., 1995. Natural Language Interfaces to Databases: An Introduction. *Journal of Natural Language Engineering, vol. 1*. Cambridge University Press.

Brill, J., et al., 2001. Data-intensive question answering. In *TREC 2001,* NIST.

Buchanan, B.G., Moore, J.D., Forsythe, D., Carenini, G., Ohlsson, S. and Banks, G., 1995. An intelligent interactive system for delivering individualized information to patients. In *Artificial Intelligence in Medicine, 7*.

Cawsey, A., Grasso, F. and Jones, R., 1999. A conversational model for health promotion on the WWW. In *AIMDM'99*, Springer-Verlag.

Cimino, J.J., Johnson, S., Aguirre A., Roderer N., and Clayton P., 1992. The Medline Button. In *16th Annual Symposium on Computer Applications in Medical Care*.

Copeland, J., 1993. Artificial Intelligence: A philosophical Introduction. Blackwell.

Craven, M., and Kumlien, J., 1999. Constructing biological knowledge bases by extracting information from text sources. In *ISMB'99, 8th International Conference on Intelligent Systems for Molecular Biology*.

Demner-Fushman, D. and Lin, J., 2005. Knowledge extraction for clinical Question Answering: preliminary results. In *AAAI'05 Workshop on Question Answering in Restricted Domains,* AAAI press.

Girju R. 2003. Automatic Detection of Causal Relations for Question Answering. In *ACL 2003 Workshop on Multilingual Summarization and Question Answering*. ACL press.

Hartrumpf, S., 2006. Adapting a Semantic Question Answering System to the Web. In *MLQA'06, EACL Workshop on Multilingual Question Answering,* ACL press.

Jacquemart, P. and Zweigenbaum, P., 2003. Towards a medical Question Answering system: a feasibility study. *Studies in Health Technology and Informatics*, IOS Press.

Jurafsky, D., and Martin, J.H., 2000. *Speech and language processing*, Prentice Hall, NJ, USA.

Laurent, D., Seguela P., and Negre, S., 2006. QA better than IR?. In *MLQA'06, EACL Workshop on Multilingual Question Answering*, ACL press.

Lin, J. and Demner-Fushman, D., 2005. "Bag of Words" is not enough for Strength of Evidence Classification. In *AMIA 2005, the Annual Symposium of the American Medical Informatics Association*.

MEDLINE database, 2006. National Library of Medicine. http://www.ncbi.nlm.nih.gov/PuBMed

McCray, A.T. and Nelson, S. J., 1995. The semantics of the UMLS knowledge sources. *Methods Inf Med 34, 1-2*.

McKeown, K., et al., 2001. PERSIVAL, a System for Personalized Search and Summarization over Multimedia Healthcare Information. In *JCDL 2001,* ACM press.

Moldovan, D., et al., 2003. LCC Tools for Question Answering. In *TREC 2003,* NIST.

Molla, D., et al., 2003. NLP for Answer Extraction in Technical Domains. In *EACL 2003,* Morgan Kaufmann.

Niu, Y., and Hirst, G., 2004. Analysis of Semantic Classes in Medical Text for Question Answering. In *ACL Workshop on Question Answering in Restricted Domains*. ACL press.

Niu, Y., Hirst, G., McArthur, G. and Rodriguez-Gianolli, P., 2003. Answering clinical questions by identifying roles in medical texts. In *ACL Workshop on Natural Language Processing in Biomedicine, 41st annual meeting of the Association for Computational Linguistics,* ACL press.

Richardson, W. S., Wilson M., Nishikawa J. and Hayward R., 1995. The well-built clinical question: a key to evidence-based decisions. In *American College of Physicians Journal Club 123, 3*.

Rinaldi, F., Dowdall, J., Schneider, G., and Persidis, A., 2004. Answering Questions in the Genomics Domain. In *ACL 2004 Workshop on Question Answering in Restricted Domains*. ACL press.

Rindflesch, T.C., Tanabe, L. Weinstein, J. N. and Hunter, L., 2000. Edgar: Extraction of drugs, genes and relations from the biomedical literature. In *Pacific Symposium on Biocomputing*.

Sackett, D., Straus, S., Richardson, S., Rosenberg, W., and Haynes, R., 2000. *Evidence-based medicine: how to practice and teach EBM*. Churcill Livingstone.

Salton G., and Buckley, C., 1988. Term weighting approaches in automatic text retrieval. *Information Processing and Management, 25, 5.*

Sekimizu, T., Park, H. and Tsujii, J., 1998. Identifying the interaction between genes and gene products based on frequently seen verbs in Medline abstracts. *Genome Informatics*, Universal Academy Press.

Sneiders, E., 2002a. Automated Question Answering Using Question Templates that Cover the Conceptual Model of the Database. In *NLDB'2002,* Springer-Verlag.

Sneiders, E., 2002b. *Automated Question Answering: Template-Based Approach*. PhD thesis, Royal Institute of Technology, Sweden.

Spasic, I., Nenadic, G., and Ananiadou, S., 2003. Using domain-specific verbs for term classification. In *ACL 2003 Workshop on Natural Language Processing in Biomedicine,* ACL press.

Teufel, S., Hatzivassiloglou, V., McKeown, K., Dunn, K., Jordan, D., Sigelman, S., and Kushniruk, A., 2001. Personalised medical article selection using patient record information. In *AMIA 2001, the Annual Symposium of the American Medical Informatics Association*.

Toulmin, S., 1958. *The uses of argument*. Cambridge University Press.

Voorhees, E., 2001. The TREC Question Answering Track. In *Natural Language Engineering 7*, Cambridge University Press.

Weizenbaum, J., 1966. Eliza – A computer program for the study of natural language communication between man and machine. *Comm. of the ACM 9*, ACM press.

CLINICAL AND TRANSLATIONAL SCIENCE INFORMATICS INFRASTRUCTURE

A Framework and Case Study

Arkalgud Ramaprasad, Annette L. Valenta
University of Illinois at Chicago, Chicago, IL 60607, USA
prasad@uic.edu, valenta@uic.edu

Ian Brooks
National Center for Supercomputing Applications, Urbana, IL 61801, USA
ian@ncsa.uiuc.edu

Keywords: Clinical and Translational Science, Informatics Infrastructure.

Abstract: This paper presents a comprehensive socio-technical framework for the design and development of a Clinical and Translational Science Informatics Infrastructure (CTSII). Based on our experience with developing and applying the framework we present a case study to illustrate the issues that arise in the creating a CTSII, and how possibly these issues can be resolved. The framework is presented as a menu with six columns, each column representing a dimension of the framework. The categories within each dimension can be concatenated, with the conjunctive phrases/words between the columns, to form sentences that describe all the functions of the CTSII. Elucidation of all the combinations will provide an exhaustive list of all the possible functions of CTSII.

1 INTRODUCTION

In 2002, the National Institutes of Health (NIH) in the US charted a roadmap for this century to identify opportunities and gaps in biomedical research in order to make the biggest impact on the progress of medical research (NIH Office of Communications, 2003). The roadmap seeks to foster new pathways to discovery, to develop innovative research teams of the future, and to reengineer the clinical research enterprise (NIH Office of Portfolio Analysis and Strategic Initiatives, 2006). It seeks to create a new discipline called Clinical and Translational Science (CTS) to reduce the time-to-practice of biomedical scientific discoveries, and the time-to-research of clinical and community health care issues (NIH Office of Portfolio Analysis and Strategic Initiatives, 2007).

Clinical and Translational Science (CTS) by definition is interdisciplinary; however, it is difficult to foster interdisciplinary research cutting across basic, animal, clinical, and public health disciplines. One barrier to such research is disciplinary silos that often manifest themselves in the form of departments, colleges, journals, and conferences. A well designed CTS Informatics Infrastructure (CTSII) can help break these barriers.

It is natural for people to know more about the research and researchers in their discipline than in others. Disciplinary research is the foundation of academic advancement, at least in the short run. The incentives systems in universities are woven around disciplinary productivity and the performance is evaluated by peers in the discipline. Consequently, the silos foster relationships within their boundaries rather than across them. While disciplinary research is necessary, it is also necessary to cut across these silos to develop CTS. How can CTSII help?

There is a disconnection between the availability and the use of informatics tools and techniques. Many popular consumer informatics tools demonstrate immense potential. Our objective is to import these tools and techniques and apply them to create an effective CTSII. Metaphorically, the ideal CTSII is a combination of Google™, Facebook™, Amazon™, and Orbitz™. It should have the global indexing, ranking, and search capabilities of Google™; the social networking capabilities of

Facebook™; the data mining, cataloging and customer [researcher] relationship management of Amazon™; and the complex scheduling [chaining] capabilities of Orbitz™. Analogues of these four systems, which have revolutionized consumer informatics, will serve as excellent bases for the design of CTSII (Valenta et al., 2007).

2 CTSII FRAMEWORK

To break the silos of research, while simultaneously advancing science, the CTSII should facilitate back and forth translation of information between basic researchers, animal researchers, clinical investigators, and public health researchers. (We use information to generically connote data, information, and knowledge.) It must support the translation of information between the sub-disciplines of each group as well (Valenta et al., 2007).

The quality and quantity of information translation will determine the effectiveness of CTS. In the following sections we present a systematic framework to analyze and design CTSII. We are currently using the framework to develop the CTSII at our university. The framework incorporates, integrates, and extends the ideas from the CTS proposals that have been funded by NIH. It has been presented to and discussed with a large group of researchers across the campus from a wide range of disciplines – including medicine, nursing, applied health sciences, engineering, business administration, public health, and pharmacy. We will discuss the issues that have been raised during our discussions, and how we plan to address them.

CTSII is not just a technological infrastructure, but also a social, psychological, organizational, and educational one – a fact that can be easily overlooked. The proposed system will, by its very design, restructure workgroups, causing stress to the organization, its social groups, and individuals. Appropriate education, consultation, training, change management, evaluation, and assessment mechanisms will be critical for the success of CTSII.

2.1 CTSII Menu

We present our CTSII framework as a menu with six columns (Figure 1), each representing a dimension of the framework. The six dimensions represent: (a) the different types of integration central to translation, (b) the different areas of research that have to be translated, (c) the resources available for translation, (d) the diseases that form the focal point of translational research, (e) the methodological steps in any research (including translation research), and (f) the tools for translation. In fact, the menu is a method of representing a matrix with six dimensions; each dimension being represented by a column. The categories within each dimension can be concatenated, with the conjunctive phrases/words between the columns, to form sentences that describe all the functions of the CTSII. Some example combinations follow:

- Lateral integration of basic research databases related to HIV/AIDS for theory construction using scientist relationship management.
- Temporal integration of public health researchers related to asthma for empirical testing using scientific workflow management.

It can be seen that even with the abbreviated list of entries in the columns, the total number of combinations is very large, indicating the complexity of CTSII. Elucidation of all the combinations will provide an exhaustive list of all the possible functions of CTSII. It would be difficult, if not impossible, to incorporate all of them in one system – they have to be prioritized. The following provides a description of the six dimensions and a sample of the categories within each.

2.1.1 Integration Dimension

Integration is one of the major driving forces behind CTS. It has been a somewhat elusive but important goal sought through earlier initiatives in interdisciplinary and multi-disciplinary research. The objective of CTS is to substitute serendipitous integration with systematic integration.

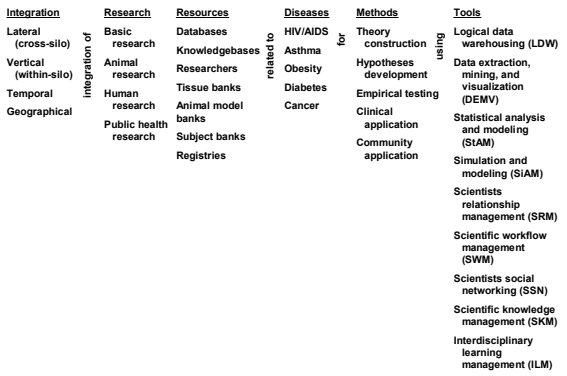


Figure 1: CTSII Menu.

The integration dimension has four categories. They are: (a) Lateral integration (cross-silo); (b) Vertical integration (within-silo); (c) Temporal integration (over time); and (d) Geographical integration (across many locations).

CTS requires integration across and within silos of basic research, animal research, human research, and public health research. To be effective, the research also has to be integrated longitudinally – over time, and across many geographical locations where the research resources may be located. Hence the four categories of the integration dimension.

Each of the four types of integration imposes a different set of requirements on the CTSII. In addition to the purchase and installation of the hardware, software, and networks, the participants will have to be informed, and trained to use the new infrastructure, and the processes of scientific collaboration will have to be reengineered to utilize the new infrastructure.

2.1.2 Research Dimension

CTS encompasses four types of research. They are: (a) Basic research; (b) Animal research; (c) Human research; and (d) Public health research.

Each of these phases includes many components; for example, the human research phase includes human trials, treatment modalities, and clinical practice. Similarly, public health research includes dissemination of the results to the public and community.

These four phases encapsulate the concept of moving basic research to the patient's bedside and the public – the central tenet of CTS. While these four phases are commonly presented as a progressive sequence, research ideas may originate in any phase and move across these phases in any order. Thus, research ideas may originate from basic research and may be fed-forward directly into human research; or, they may originate in public health research and may be fed-back directly to animal research.

One of the major concerns of CTS is that each phase tends to be a silo. These silos are reinforced by norms of the disciplines and associated incentives. The silos inhibit feed-forward and feed-back. Consequently, the movement across the phases tends to be slow and not smooth. A significant body of research may accumulate in a phase without any impact on the subsequent phases through feed-forward or on prior phases through feed-back. When this happens, both the creative and corrective value of feed-forward and feed-back is lost. Streamlining the feed-back and feed-forward mechanisms using CTSII, on the other hand, can improve both the efficiency and effectiveness of translation (Ramaprasad, 1979, 1982, 1983). Similarly, silos within silos can inhibit feed-in. Streamlining feed-in using CTSII can lead to improvement in the quality of feed-forward and feed-back.

2.2 Resources Dimension

CTS requires integration of a large number of resources. They are: (a) Databases – central, homegrown, relational, flat files, etc.; (b) Knowledgebases – structured, unstructured, text, formal, informal, etc.; (c) Researcher databases – directories, résumés, profiles, etc.; (d) Tissue banks; (d) Animal model banks; (e) Subject banks – deidentified subjects, identified subjects, volunteers, etc.; and (f) Registries

Under each of the above categories, there is likely to be a large number of subcategories, and ultimately a larger number of actual resources. Developing an inventory of these resources will be a key step in developing the informatics infrastructure.

Researchers often focus exclusively on the integration of databases as a requirement of the CTSII. While databases are important, integration of information about other resources is equally important. A clinical researcher, for example, probably does not need access to the genomic database used by a basic researcher to discover a gene marker for breast cancer, but needs information about the marker and how to test for it. The clinical and basic researchers need to know which other researchers can help them develop a reliable test for the marker, how they can obtain a panel of subjects for a trial, and the tools to evaluate the results of the trial.

The CTSII should, ideally, replace the usually ad-hoc processes for accessing these resources with more efficient and effective processes. By making the resources visible and accessible to all, the CTSII should improve both the quality and utilization of these resources.

2.3 Diseases Dimension

Different diseases may require different combinations of capabilities in the CTSII. While gene-based diseases may require the ability to handle large volumes of genomic data, environmentally induced diseases may require the ability to manage disparate public health data. Similarly, some diseases may require the ability to

educate and interact with the public health workers and the community physicians and nurses.

The CTS in an institution may be focused on a few or a large number of diseases. Following is an illustrative list of diseases suited to the current research at our university: HIV/AIDS, asthma, obesity, diabetes, and cancer.

Focusing on the above will fit our university's expertise as well as its mission as a premier urban research university. Other institutions may have other priorities based on their vision, mission, strategy, and environment. The design of the CTSII will naturally depend upon the different informatics required for the management of the diseases in question.

2.4 Methods Dimension

The informatics requirements of different stages of research are different. The CTSII should support all the stages.

The first three categories in the methods dimension are standard stages in research methodology. Carrying them forward to clinical and community application, the last two categories, are an essential part of translation. Thus, the five categories are: (a) Theory construction; (b) Hypotheses development; (c) Empirical testing; (d) Clinical application; and (e) Community application.

The type of translation required for translational theory construction may be quite different from that required for translational clinical application. The former may require metaphorical translation from one discipline to another; the latter may require a methodological translation. (Please see section 3 below for a more detailed discussion of the different types of translation.) Thus, for the metaphorical translation, the CTSII may have to facilitate the social networking of theoreticians from the different disciplines, while for methodological translation, the methodologists from the disciplines may have to be brought together. These groups, in turn, may need access to different types of knowledge.

2.5 Tools Dimension

CTS requires integration using many tools. This dimension articulates the metaphor we used earlier for CTSII as a combination of Google™, Facebook™, Amazon™, and Orbitz™. Under each of the following categories, there is likely to be a large number of tools. In fact, many tools may span multiple categories. Developing an inventory of these tools will be a key step in developing the informatics infrastructure. The categories of tools include:

- Logical data warehousing tools;
- Data extraction, mining, and visualization tools;
- Statistical analysis and modeling tools;
- Simulation and modeling tools;
- Scientist relationship management tools;
- Scientific workflow management tools;
- Scientist social networking tools;
- Scientific knowledge management tools; and
- Interdisciplinary learning management tools.

The above is not an exhaustive list of the types of tools. There are many other types of tools, and many more are likely to emerge in the future. There will also be many more tools within each type. The difficulty is not one of the availability of tools, but of their application to CTS and in developing workflow management capabilities to integrate the tools. The cross-fertilization of the application of these tools across traditional CTS disciplines, and between non-CTS disciplines (for example: marketing, production and operations management, semiotics, computer science, and library science) will be facilitated by the CTSII.

3 CTSII CASE STUDY

In this section, we will present the key issues that have arisen as we have tried to adopt and apply the CTSII framework described above over the past year. These issues are unlikely to be unique to our institution. It is intrinsic to the nature of transformation that CTS is trying to engender. In addition to illuminating the process of application, the case also highlights the importance of considering the psychological, social, organizational, and educational aspects of the CTSII.

3.1 Not Just Databases

To many, the term informatics appears to connote only databases. A number of early meetings focused exclusively on developing an inventory of the databases and making them easily accessible to other researchers. Perhaps it reflected the participants' primary concern with their research. It took many meetings to convince the participants (a) that databases were only one type of informatics resources, and (b) that informatics should focus on a broader range of functions than simply creating, integrating, and providing access to the databases.

3.2 Oracle™ and Google™ Mindsets

Related to equating informatics with databases, there was a strong tendency to think exclusively in terms of structured databases: simple flat Excel files and more complex relational databases. This could be called an 'Oracle™' mindset to distinguish it from the 'Google™' mindset – storage and search based on unstructured repositories and documents. Part of the bias appears to be due to lack of knowledge of today's information systems' ability to index, search, and access large volumes of unstructured data in documents and other sources. Even though researchers used Google™ and similar search engines, few have knowledge of how search engines work. The bias was compounded by an unrealistic equating of the cost of shrink-wrapped databases with the cost of building a structured database for unstructured data.

3.3 Thinking Outside the SiloS

Often, informatics requirements presented by the other core groups involved in the CTS effort appear to be focused on research within their silos, and not across silos. They do not seem to be focused on the truly translational processes of feed-forward and feed-back, but focused on feed-in. While many of the informatics requirements will no doubt facilitate research, it is not clear how they will facilitate translational research. It will perhaps take some more time for the researchers to change their framework through education, consultation, and experience with the proposed CTSII.

3.4 Foundation and Frontier Requirements

A consequence of the issues discussed in the above sections is that most of the requirements presented tend to be what we have called 'foundation' requirements; for example: web-accessible databases and an interactive directory of researchers as a baseline requirement. They are necessary for CTS, and for that matter, any research. They are unlikely, however, to transform the CTS research at the institution and to distinguish one institution from another. That requires the 'frontier' capabilities. These capabilities reflect some of the best practices across different types of organizations and may have to be adapted to CTS. While many of the researchers are aware of these capabilities and have used them, they do not often see their application to CTS. The barriers to the transfer of these best practices are many, one being the inaccurately perceived high cost of performing certain functions. The cost of full-text search is an example of the incorrect perception. The complexity and difficulty of social network analysis is another example of such perception.

It will likely be a while before the researchers will start using CTSII as a tool for CTS research instead of just a service to improve their current research. Until then, the stated requirements are likely to be at the foundation end of the spectrum rather than at the frontier end. Movement of the researchers' thinking along the spectrum will be an important part of the transformation. When, in fact, they move into the frontier, there will likely be a sudden cascade of new and innovative requirements of CTSII.

3.5 Metaphoric Success

The metaphor 'CTSII = Google™ + Facebook™ + Amazon™ + Orbitz™' was very successful in communicating our ideas and to reframe their thinking. Many people, given their age, were not familiar with Facebook™, although some reported their children were using it or a similar system. They were less familiar with Orbitz™ than with Expedia™ and Travelocity™. We chose Orbitz™ as a result of the recent experience of one of the authors with booking a complex international trip. Of course, most researchers were familiar with and had used Google™ and Amazon™.

Almost everybody liked the metaphor once it was explained to them. Many, especially those in informatics related disciplines, intuitively grasped its meaning, and immediately liked it. We still had to explain the application of the metaphor to the design of CTSII, however. The barriers to application of the metaphor were similar to those explained in 3.3 above. The metaphors and the menu have been the anchors of all the recent discussions on CTSII.

3.6 Feed-forward, Feed-backward, and Feed-in

In our eagerness to adhere to the spirit and letter of CTS, we initially proposed feed-forward and feed-back mechanisms as the bases for translation between the four types of research. Feed-in was added in response the expressed need of improving the informatics within each discipline, too; however, the feed-concepts did not appear to resonate with the CTS researchers. As a consequence, we renamed "feed-forward and feed-back" to "two-way

translation of information". We continue to believe, however, that the feed-concepts are central to CTS, and part of the transformation will be to understand and apply these concepts systematically.

In a sense, the CTS paradigm has close parallels to the Continuous Improvement Paradigm that emerged in the context of the Total Quality Management (TQM) movement (Söderholm, 2004). The feed-concepts will likely be as critical to CTS as they have been to Continuous Improvement.

3.7 Technological Bias

Given the focus of most researchers on the technical aspects of informatics – the databases and data warehouses, it took time to incorporate the socio-technical view in the discussions. Clearly, help desk, consultation, education, training, and other user services must be part of the CTSII. The idea of little or no human intervention in this transformation is not realistic.

3.8 Local and Global Transformation

The major impetus for CTS is to transform research. Some of this transformation can be in the CTSII, and some using CTSII. The CTSII transformation can be local or global. Local transformation is one which is innovative within an institution, but not globally. Other institutions may have adopted the innovations earlier. Global transformation is one which is innovative within the institution and without. It is the first of a kind, anywhere.

Both local and global CTSII transformations are necessary for the success of CTS efforts in an institution. Many local transformations may be necessary to bring an institution on par with other institutions; at least a few global transformations will be necessary to provide a competitive advantage to the institution.

4 CONCLUSIONS

An informatics infrastructure for clinical and translational science (CTS) can be complex. We conceptualize the information flows for CTS as a bidirectional flow of feed-forward and feed-back between basic researchers, animal researchers, clinical researchers, and public health researchers, and feed-in among researchers within a discipline. We present a six-dimensional framework as a menu to deconstruct the complexity and help specify the requirements of the clinical and translational science informatics infrastructure (CTSII) for an institution. The framework provides a simple concise way of enumerating all the functions required of a CTSII. Last, we present a case study of our experience in using the framework at our institution. The case study illustrates some of the barriers to the application of the framework and how these barriers can be overcome.

REFERENCES

NIH Office of Communications. NIH News: NIH announces strategy to accelerate medical research progress. September 30, 2003. Available at http://www.nih.gov/news/pr/sep2003/od-30.htm. Accessed May 31, 2007.

NIH Office of Portfolio Analysis and Strategic Initiatives. Re-engineering the clinical research enterprise. April 7, 2006. Available at http://nihroadmap.nih.gov/clinicalresearch/. Accessed May 31, 2007.

NIH Office of Portfolio Analysis and Strategic Initiatives. Re-engineering the clinical research enterprise. March 1, 2007. Available at http://nihroadmap.nih.gov/clinicalresearch/overview-translational.asp Accessed May 31, 2007.

Ramaprasad, A., 1979. The Role of Feedback in Organizational Change: A Review and Redefinition. *Cybernetica*, 22 (2), 105-113.

Ramaprasad, A., 1982. Revolutionary Change and Strategic Management. *Behavioral Science*, 27 (4), 387-392.

Ramaprasad, A., 1983. On the Definition of Feedback, *Behavioral Science*, 28 (1), 4-13.

Ramaprasad, A. and Rai, A., 1996. Envisioning Management of Information. *Omega: The International Journal of Management Science*, 24(2), 179-193.

Söderholm, P., 2004. Continuous Improvements of Complex Technical Systems: a Theoretical Quality Management Framework Supported by Requirements Management and Health Management. *Total Quality Management*, Vol. 15, No. 4, 511–525.

Valenta, A.L., Brooks, I., Laureto, R.A., and Ramaprasad, A., 2007. Informatics to Support Clinical and Translational Science. Journal of Health Information Management (forthcoming).

EXPLOITING SERVICE ORIENTED ARCHITECTURES FOR THE DESIGN OF E-HEALTH SYSTEMS

Marcos Da Silveira and Nicolas Guelfi
Laboratory for Advanced Software Systems, University of Luxembourg, 6, rue Coudenhove-Kalergi, L-1359 Luxembourg
marcos.dasilveira@uni.lu, nicolas.guelfi@uni.lu

Keywords: Service oriented architecture, interoperability, standards, e-health, information technology.

Abstract: The design of e-Health systems is a hard task since their requirements are complex and heterogeneous. These systems merge functional requirements with an important set of non–functional requirement like security, safety, standardization or technology related constraints concerning the hardware and software components to be used. The ICT research community has proposed recently architectural models for the development of open and dynamic distributed systems centered on the concept of "service". This approach has been followed by all the major actors of ICT for their frameworks due to its adaptation to the World-Wide-Web. This paper is a position paper where we analyze the current status and needs for e-Health systems and the limitation upon them. We present the main characteristics of middlewares that follow a service-oriented architecture and we explain how these frameworks could be exploited, as a vision to the future, to design e-health systems for a better insurance of their functional and non-functional requirements.

1 INTRODUCTION

Healthcare market is a typical example of service sector that has successfully increased its participation in the world's economy. Because of the importance of services market, in particular business and healthcare services, their development and their delivering have become the center of discussion in many levels of the society. Politicians and lowers have looking for adapted legislations and controls (Garner, 2004). Educators and industries search convenient methods to introduce *service science* (Chesbrough & Spohrer, 2006) into the continue education and to use it to prepare a new generation of healthcare professionals. A generation that needs interdisciplinary knowledge to increase the productivity and innovation of service delivering.

In a digital society, the concept of services evolves and words like e-service and web-service gains their place in the market. In health domain, these new kind of services have facing many difficulties (resistance to change habitudes, misinformation, technological phobias, etc.). However, caregivers are slowly getting used with the idea of adopting informatics devices to improve the quality of their services. Some countries are catalyzing this process forcing caregivers to use Information and Communication Technologies (ICT) to interact with public administrators (for example, social security administration and public hospitals administrations) (Currie & Guah, 2006),(GAO, 2005).

Clearly, the association of informatics with healthcare domains brings up many advantages. Healthcare industry is positioned to beneficiate of the advancements in technology and connectivity. High technological devices and software are available for healthcare services' providers and are adapted to the patient needs (patient-centered systems). Customers of these technologies can expect to achieve greater performance, reduce costs and improve patient care. Consequently, they are expanding marketshare and are pushing the transition to a digital era of *e-Health*. In this paper, the word e-Health expresses the association of e-technologies with healthcare services. Similar words have been used in the literature to express the association of specific informatics technology with healthcare, as well as telecare, telemedicine, telehealth, bioinformatics, etc. (Norris, 2002). The definition that we consider the most appropriated to our work was proposed by (Eysenbach, 2001).

"e-Health is an emerging field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the Internet and related technologies. In a broader sense, the term characterizes not only a technical

development, but also a state-of-mind, a way of thinking, an attitude, and a commitment for networked, global thinking, to improve health care locally, regionally, and worldwide by using information and communication technology"

The complexity associated with the design of e-Health systems and its application environment has driven developers to look for methodologies that can simplify their work. The service-oriented architecture (SOA) is a solution that is captivating many important industrials. SOA describes the practices and frameworks that enable the functionality and information of applications to be provided and consumed as service interfaces. In this formalism, services can be seen as network addressable entities with a well-defined, easy-to-use and standardized interface. From a technical viewpoint, SOA is essentially a collection of software services communicating with each other over a network to pass data or to coordinate activities. The software services can be implemented as gray boxes where implementation details (as well as the technologies used) are not necessarily accessible by the clients. The main expected benefit of SOA is their capacity of satisfying the requirements of complex and heterogeneous information systems domains such as healthcare. Further, services can be combined in different ways to support the evolution of processes and organizational models, for example towards virtual organizations.

The work developed at LASSY (Laboratory for Advanced Software Systems) aims to contribute to the improvement of the organizational structure of health systems proposing a SOA adapted to e-Health requirements. In this paper we identify the current status and needs for e-Health systems and the limitations upon them. Section 2 introduces some problems observed in the Healthcare structure. Section 3 presents the mains concepts of service oriented architecture (SOA) and a summary of TAPAS project. Section 4 describes the potential application of SOA for designing e-Health systems as the objectives and perspectives of RESIST project.

2 IDENTIFYING PROBLEMS IN HEALTHCARE SYSTEMS

The current situation of healthcare domain points to an excessive individualization of activities and a lack of communication inside of the institutions. Information generated to/by one caregiver is rarely shared and finishes in an archive (electronic records or not) often inaccessible by other caregivers and by the patient. The same problem had been faced in other professional activities, the most representative are the financial activities. Continuous investments on ICT solutions contribute to reduce the consequences of this problem. (Luftman et al., 1993) highlights that the effective and efficient utilization of information technology requires the alignment of IT strategies with the business ones, what was not successfully done in the past with traditional approaches. The same idea can be applied in healthcare systems waiting for attaining efficient and responsible use of ICT to solve the exigencies of healthcare centers' administrators, caregivers, patients and governments. This idea is not completely new, some hospitals use ICT in their organization from the lasts 25 years. But this initiatives are rare and do not represent the majority of the cases. According to the Department of Health and Human Service of USA (HHS), healthcare is the largest sector of the economy that has not fully embraced information technology. The Medical Group Management Association reported that only 31% of physician group practices use fully operational Electronic Health Records (EHRs). The Healthcare Information and Management Systems Society reported that 19% of hospitals use fully operational EHRs (GAO, 2005). The results of these initiatives were reported in (Currie & Guah, 2006), (Wears et al., 2006), (Keizer & Ammenwerth, 2007), but we are only at the beginning of great changes produced by the association of information technologies, network managed organizations and specialization with the healthcare domain. The current challenge is to define how to integrate the set of existing specific solutions?

Many solutions use proprietary data exchange format and do not interact with the others. Interoperability becomes now fundamental to healthcare communities. Industries, research groups and governments are investing time and money to solve this problem. Many standards are emerging of these collaborations, but few of them are currently integrated in most commercial systems (HL7, 2007), (CEN, 2006), (DICOM, 2003). Unclear legislation, lucrative market, implementation costs, physicians' reluctance and disorganization contribute to prolong that reality. The way that IT solutions have been implemented in healthcare systems are not exactly what users expect and it will not solve all problems in this domain, but it is a beginning to improve the healthcare services' quality. The benefits that can be expected are:

- For the consumer: Higher quality care; Reduction in medical errors; Fewer duplicate treatments and tests; Decrease in paperwork; Lower healthcare costs; Constant access to

health information; Expansion of access to affordable care;

- For the public health sector: Early detection of infectious disease outbreaks around the country; Improved tracking of chronic disease management; Ability to gather de-identified data for research purposes; Evaluation of health care based on value, enabled by the collection of price and quality information that can be compared.

All these benefits contribute to improve the life quality of patients, but our main motivation is in reducing (or eliminating, if possible) the alarming number of deaths related to medical error that could be prevented if an efficient IT system was implemented.

The implementation of a good IT requires the selection of an architecture that can represent as real as possible the way of working of stakeholder in healthcare domain. The next section presents the main characteristics of the architecture that we consider as the most appropriated to e-Health systems.

3 SERVICE ORIENTED ARCHITECTURE

The concept of service-oriented architecture is not new, but with the advent of recent platform-independent programs and platform-neutral data models, this architecture took more attention. A unique formal definition of SOA does not exist, and discussion about it is out of the scope of this paper. However, we can highlight some important aspects (Gupta, 2007):

- The service-consumer needs must be specified independently of the service-provider component. A reasoning engine takes the responsibility of aligning the two specifications. This engine, often named Assembler component, bridges the gap between the two end-elements (consumer and provider). (Loose Coupling)
- The technology used by the service-consumer is completely independent of the one used by the service-provider and vice-versa. (Platform Independence)
- The communication between the two end-elements must not be dependent of the protocol. A variety of communication protocol must be available to request/offer a Service. The choice of protocol must not modify the quality of the service. Binding to a specific protocol must take place at run-time/deployment-time, and not at the design or development time. (Communication Protocol Independence)
- Each of the provider- and consumer-service components should be able to be implemented in their own time, according to their own life-cycle. (Life-cycle Independence)

Specially designed for flexibility and reuse, a SOA enables organizations to easily integrate systems, data, applications, and processes through the linking of services. SOA was first proposed to satisfy the necessities of business design and implementation that was not supplied by distributed-oriented architecture (DOA, e.g. CORBA, J2EE, etc.). It can be used as a complementary platform or to substitute DOA for providing intra and inter-business services. The general concept of SOA is close of the one of DOA, however there are small differences and the sum of these features leads to a radically different set of properties for enterprise-level system modeling and design (Baker & Dobson, 2005). SOA is a software architecture that uses loosely coupled software services to support the requirements of business processes and software users. It is technology neutral and has no standard specification of the components interface, what offers dynamism and flexibility to the system. The utilization of asynchronous message exchange instead of function calls allows services to be executed without share details about its implementation or platform. SOA enables uncomplicated connectivity by abstracting dependencies away from each application into an e-service (or other brokering service, e.g. web-service). Applications can then easily be connected to the broker through modular components. And each application can be modified whenever necessary to support flexible and dynamic business processes through platform-independent.

The design and development of a SOA implies a different way of organizing the system. Designers must to change the component oriented vision and think about services' "properties and attributes". It requires the application of a number of techniques – stemming from disciplines such as Enterprise Architecture, Business Process Modeling, Component-based development and Object-oriented methods – to produce modular, reusable and replaceable software applications.

It seems to be more complicated then DOA. In fact, it demands an extra effort on coordination and interface definition to understand information assets and link to business process. However, once created the basic structure, modifications in a service does not request the understanding of the whole model.

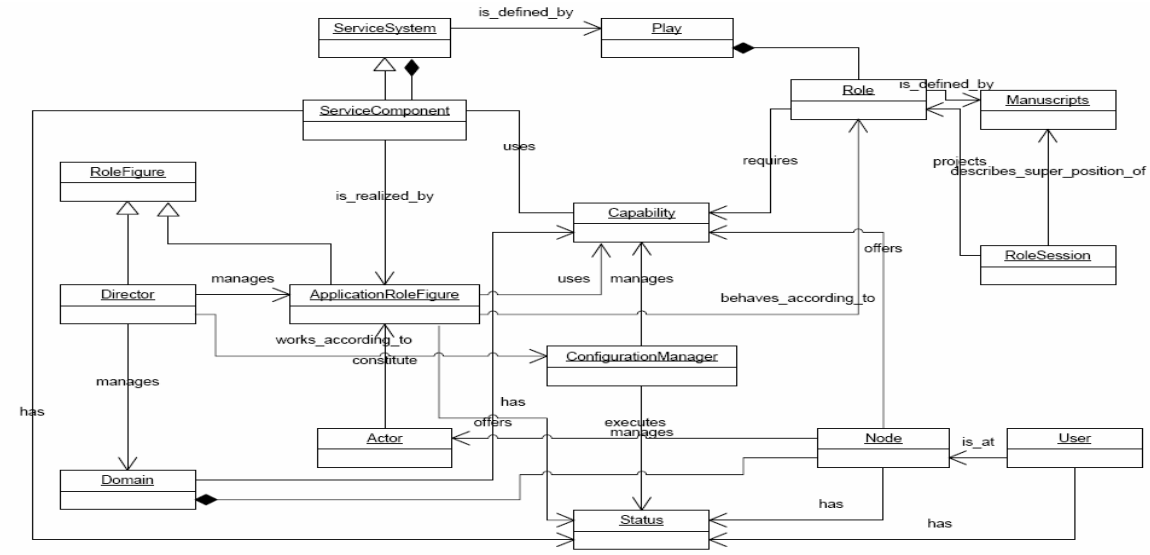


Figure 1: TAPAS basic architecture.

Services can be seen as gray boxes that have a description file each one (normally in XML) which specifies the interface and some other attributes of the service. Further, most gray boxes in a SOA are represented as legacy applications. However these will usually need re-engineering to provide access to their functionality and data from a wider population of consumers. The key of a good SOA implementation depends on the specification of the services' interface and the reasoning engine to mach requested services with offered ones. An usual way to solve this problem is to implement different layers of functionalities that see the services with different granularity. The approach described in TAPAS project (TAPAS, 2007), for example, proposes three layers: *Service, Play* and *Network*. The service layer consists of service components definition and the network layer consists of nodes specification (places where services can be executed). The play layer is used to connect consumers with providers and to identify the node to execute the service.

The next section presents more details about TAPAS' concepts, but a complete information about this approach can be found in the project website (TAPAS, 2007).

3.1 TAPAS' Main Concepts

The TAPAS project is a research project developed at NTNU which aims at developing and architecturing for network-based service systems with: a) flexibility and adaptability; b) robustness and survivability; and c) QoS awareness and resource control. The goal is to enhance the flexibility, efficiency and simplicity of system installation, deployment, operation, management and maintenance by enabling dynamic configuration of network components and network-based service functionality.

The TAPAS basic architecture, illustrated in Figure 1, is founded on generic *Actors* (software components) and on the *Nodes* of the network. The actors play *Roles* defined in a *Manuscript*, which consists of an EFSM (Extended Finite State Machine) extended with rule-based policies. The nodes may be servers, routers, switches, and user terminals, such as telephones, laptops, PCs, PDAs, etc. The idea is to associate the design model to a theatre organization. The Actors are coordinated by Director that also represents a Domain. Service System consists of a set of Service Components, which are units related to some well-defined functionality defined by a Play. A Play represents several Actors playing different Roles. The ability of Actors to play Roles depends on the defined required Capability and the matching offered Capability in a Node where they intend to execute.

Capability is an inherent property of a Node or user, it may be resources (e.g. CPU, hard disk, transmission channels), functions (e.g. printing, encryption devices), or data (e.g. user login, access rights) (Lühr, 2004). Where and how Actors are installed and roles executed are decided by the *Configuration Manager*. It is responsible for obtaining a snapshot of all system resources, and taking decisions about the configuration. This brief introduction to TAPAS gives an idea of the potentiality of this approach. The next section explains how TAPAS can be applied into healthcare domain.

4 EXPLOITING SOA IN E-HEALTH SYSTEMS

Analyzing the weaknesses of healthcare systems' organization described in section 2 and the features that can be expected from SOA, we can imagine that SOA is an interesting approach to supply many identified problems of health sector. In this architecture we can identify services associated with the specification: of different classes of users; of polices to increase the security and the reliability; of interfaces to allow communication between applications and between systems; of common terminologies to facilitate information exchange and information understanding; etc.

Composed of a set of (reusable) services, this structure facilitates the management of the offered services and allows to measure more accurately the quality of those services (QoS). Improvement of QoS is an important requirement of e-health systems. Thus, in this structure the patients that can input data to accurate the QoS (having a proactive role) and consequently trust more on the (managers) selection of actors for the required services according to the level of QoS specified (feedback). Further, the healthcare administration centers can obtain all (or much more than today) necessary information to take decisions about the service. In a sense, the services-providers (caregivers, equipments and software's providers, pharmacies, etc.) are pushed to improve their QoS in order to keep their marketshare.

Following this trends, TAPAS proposes to create a plug-and-play environment where services and users can be dynamically included/logged on, excluded/logged off or modified (moving, changing of resources, changing the quality of services, etc.) without a big effort. Initially developed to general propositions, the current version of TAPAS is not completely adapted to healthcare systems. The conception and implementation of a TAPAS top level to include specific properties of healthcare sector is one of the objectives of RESIST research project, in development at University of Luxembourg, LASSY (Laboratory for Advanced Software SYstems) workgroup.

Evaluating TAPAS properties we highlight the well defined distributed management structure, separating the operational specification from the management specification. Where decisions are distributed and decision-makers can have different priorities and criteria. These elements together with the user-friendly description of the operational architecture (theatre metaphor) facilitate the complex definition of the architecture lifecycle of a healthcare organization. In other words, TAPAS assists the design and implementation of the following tasks: modeling, service composition, deployment and execution management.

TAPAS is structured clearly and formally. The description of the *Role* by means of *Manuscripts* in XML files follows the international trends to improve the syntactic interoperability and standardize business interface. RESIST project plan to use this facilities to benefit from ontological standards (SNOMED, 2007), when described in XML. The rule-based reasoning engine proposed into this architecture is an intelligent solution to associate dynamically services requests with services providers and to guarantee that the specified quality threshold will be respected if there is an *actor* that can satisfy it. In the specification of TAPAS we can observe that the authors also introduce tools to detect, diagnosis and recovering faults. These tools are in development, but it can become an interesting alternative to improve security, reliability and availability properties. This is an innovation that we can not be found in commercial tools whose base their security only on the communication/data access security policies.

5 DISCUSSION AND CONCLUSIONS

If the gap between public knowledge and academic curriculum isn't large enough, the gap between academia and professional practice is a gaping hole. While academic departments insists on doctor-centered teachings in an ideal organizational environment, medical industries have innovating proposing adapted solutions to specialized needs. Some existing medical equipments can be easily manipulate by ten-year-old boys and can give precise information about the health state of the patient. This is one of the contributions of e-Health. Moving away from classical methods to focus on delivering quality services that are in-line with the new context of modern medicine is the challenge of the next decades. Informatics came to give support to it in order by proposing tools to mange, integrate, disseminate and generate health information.

Proving its worth in different business domain, ICT solutions have been implemented on many hospitals, leading to reduce medical errors and to improve the quality of medical services. But the complexity of managing all needs of stakeholders in health domain had carried out many difficulties on e-Health implementations.

The solution proposed in this paper is the designing and implementation of e-Health systems

based on SOA definitions. We remind that in this architecture, core business capabilities are encapsulated within independent software services, and these services are leveraged by various front-end applications to fulfill business requirements. The main properties of this approach are the use of business-oriented services; message-based interactions with "gray box" implementations; communication over a network; platform neutrality; service description and discovery; and loose coupling between system components (Kawamoto & Lobach, 2007). This set of properties leads to a simpler approach to software design and implementation and to enhancing re-use of existing IT resources. What gives to the system the ability to adapt to changing business requirements in a flexible, agile manner; and the potential for significant cost saving.

Lessons learned from countries' pioneer for the e-Health implementation shows that interoperability is a critical and crucial aspect for any national (and international) program. Many researches are contributing to develop it, but data exchange between softwares and between systems (respecting international directive for data control) still is a big challenge. These lessons had been taken into account during the development of RESIST project. The key idea is to improve the life quality of patients in Luxembourg, proposing a dynamic, flexible and standardized infrastructure to integrate healthcare application. The Plug-and-Play middleware developed in TAPAS project seams to be a good option to reach the RESIST aims. Our contribution in short term is the specification of an architectural framework to cover the stakeholders' necessities and the implementation of a top level in TAPAS middleware to adapt it to healthcare applications' needs. A case study has been specified to implement home monitoring in cardiology's departments. Current works targeting the identification of the requirements to create a Luxembourgish e-health environment. It deals with patient interests, caregivers' capabilities, healthcare centers' managements and governmental regulations and administration.

REFERENCES

Baker S.; Dobson S.: Comparing service-oriented and distributed object architectures. In (Robert Meersman and Zahir Tari et al, editors): *Proceedings of the International Symposium on Distributed Objects and Applications*, LNCS. Springer Verlag, 2005.

CEN/TC 251: European Standardization of Health Informatics, 2006. Retrieved June 10, 2007, from http://www.centc251.org/

Currie W. L.; Guah M. W.: IT-enabled healthcare delivery: *The U.K. National Health Service. Information Systems Management*, 2006, Vol. 23, Issue 2, pages 7-22.

Chesbrough H.; Spohrer J.: A research manifesto for services science. *Communications of the ACM,* July, 2006, Vol. 49, Issue 7, pages 35-40.

DICOM: Part 1: Introduction and Overview. Retrieved June 10, 2007, from http://medical.nema.org/dicom/2003/03_01PU.PDF

Eysenbach G.: What is e-health? *Journal of Medical Internet Research*, Apr–Jun, 2001, Vol. 3, Issue 2: e20.

Garner C. A.: Offshoring in service sector: Economic impacts and policy issues. *Online paper,* 2004. Website of Kansas City Bank. http://www.kansascityfed.org/Publicat/Econrev/PDF/3Q04Garn.pdf

GAO - United States Government Accountability Office: Health information technology: HHS is taking steps to develop a national strategy. *Report to the Chairman, Committee on the Budget, House of Representatives*, May, 2005. Retrieved June 10, 2007, from http://www.gao.gov/new.items/d05628.pdf

Gupta S.: Service Oriented Architecture - Part 2. Java Boutique. Retrieved June 10, 2007, from http://javaboutique.internet.com/tutorials/serv_orient/index-4.html

Health Level Seven. Retrieved June 10, 2007, from http://www.hl7.org/

De Keizer N. F.; Ammenwerth E.: The quality of evidence in health informatics: How did the quality of healthcare IT evaluation publications develop from 1982 to 2005?. *International Journal of Medical Informatics*, in press, 2007.

Kawamoto K, Lobach D.: 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, Vol. 14, pages 146 –155.

Luftman J. N., Lewis P. R., Oldach S. H.: Transforming the enterprise: The alignment of business and information technology strategies. *IBM Systems Journal*, 1993, Vol. 32, Issue 1, pages 198 – 221.

Lühr E.: Mobility support for wireless devices - within the TAPAS platform. *Master thesis at Norwegian University of Science and Technology* – NTNU, 2004.

Norris, A.C.: Essentials of telemedicine and telecare. *John Wiley&Sons*, 2002

Systematized Nomenclature of Medicine for Clinical Terms, SNOMED-CT international release 2007. Retrieved June 10, 2007, from http://www.snomed.org/snomedct/standards.html

TAPAS. Retrieved June 10, 2007, from http://tapas.item.ntnu.no/wiki/index.php/Main_Page

Wears R. L.; Cook R. I.; Perry S. J.: Automation, interaction, complexity, and failure: A case study. *Reliability Engineering and System Safety*, 2006, Vol. 91, pages 1494–1501

ASYMS©-SERAT: A SIDE-EFFECT RISK ASSESSMENT TOOL TO PREDICT CHEMOTHERAPY RELATED TOXICITY IN PATIENTS WITH CANCER RECEIVING CHEMOTHERAPY

Julie Cowie, Kevin Swingler, Clare Leadbetter
Department of Computing Science and Mathematics, University of Stirling, Stirling, FK9 4LA, UK
jco@cs.stir.ac.uk, kms@cs.stir.ac.uk, cll@cs.stir.ac.uk

Roma Maguire, Kathryn McCall, Nora Kearney
Cancer Care Research Centre, Department of Nursing and Midwifery, University of Stirling, Scion House
Stirling Innovation Park, Stirling, FK9 4NF, UK
roma.maguire@stir.ac.uk, kathryn.mccall@stir.ac.uk, nora.kearney@stir.ac.uk

Keywords: Risk modelling, side-effect prediction, cancer chemotherapy.

Abstract: Patients undergoing chemotherapy want specific information on potential toxicities of their treatment. Such information includes what side-effects they are likely to experience, how severe these side-effects will be, how long they will experience them for, and the best ways of managing them. As well as improving the experiences of patients, information about potential side-effects may also be of significant benefit clinically, as patients who are 'at risk' of developing certain toxicities may be identified, facilitating more targeted, cost-effective interventions. This paper describes research that uses risk-modelling techniques for identifying patterns in patient side-effect data to aid in predicting side-effects patients are likely to experience. Through analysis of patient data, a patient can receive information specific to the symptoms they are likely to experience. A user-friendly software tool ASyMS©-SERAT (Advanced Symptom Management System-Side-Effect Risk Assessment Tool) has been developed, which presents side-effect information to the patients both at the start of treatment and reviews and monitors predictions with each new cycle of chemotherapy received.

1 INTRODUCTION

In this paper we discuss the development of a user-friendly software tool ASyMS©-SERAT (Advanced Symptom Management System-Side-Effect Risk Assessment Tool), designed to provide patients undergoing chemotherapy treatment with a personalised prediction of possible side-effects they are likely to experience. Prediction of possible toxicities is achieved through the use of risk modelling techniques, which facilitate a better understanding of how a patient's personal information, the chemotherapy regime they are undertaking, and any previously experienced symptoms (if appropriate) contribute to the likelihood of future symptoms occurring.

Section 2 provides the reader with an overview of the project; detailing the current state of symptom modelling, risk modelling and how the approach can be applied to prediction modelling. The methodology adopted is also discussed as well as the actual aims of our research. In Section 3 we present the software ASyMS©-SERAT, discuss possible ways in which the system might be used and show the type of prediction information which can be provided. We conclude with Section 4, discussing work-to-date and ways in which we hope to develop ASyMS©-SERAT in the future.

2 PROJECT BACKGROUND

2.1 Symptom Modelling

In the UK approximately 277,000 individuals are diagnosed with cancer each year (CRUK, 2003) and this figure is projected to significantly increase over the next decade (SEHD, 2005). A majority of these

individuals are likely to receive chemotherapy treatment at some stage of their illness. The toxic effects of chemotherapy puts patients at risk of developing a number of side-effects, some of which can become serious and life threatening if not detected and managed early. Approximately 9-21% of patients receiving chemotherapy are hospitalised due to such severe treatment related toxicity (Chen-Hardee et al. 2006; Du, Osborne, & Goodwin 2002;Kuderer et al. 2006; Polednak 2004) and 10% of patients die as a result of them (Kuderer, Dale, Crawford, Cosler, & Lyman 2006).

The effective monitoring and management of symptoms in this patient group is vital. However, it is now recognised that symptoms in patients with cancer are often poorly assessed and managed (National Institute for Health, 2002). Factors such as inadequate patient provider communication (Cleeland et al, 1986), and poor symptom assessment (Cleeland et al, 1994) have been cited as being contributory factors. The recent changes to the organisation of cancer services may also contribute to the sub optimal management of symptoms. With the focus of care now being in the home and out-patient setting, patients are left to manage the majority of side-effects on their own without direct supervision from health care professionals; this may leave them feeling anxious and having lack of control over their illness and treatment (McCaughan & Thomson, 2000). Furthermore, patients with cancer often find the unpredictability and diversity of potential side effects difficult to deal with (Cohn 1982; Tierney, Taylor, & Closs 1992).

Patient education is therefore fundamental to effective symptom control. It is widely acknowledged that patients with cancer want information on how to manage the symptoms and side effects associated with their disease and treatment (McCaughan & Thompson 2000; Skalla, Bakitas, Furstenberg, Ahles, & Henderson 2004). However, they often report feeling overloaded with the wealth of information provided and as a result experience problems with retaining and retrieving it (Skalla et al. 2004). As a consequence, there have been calls for the provision of information on cancer therapies, which is tailored to patients' individual characteristics and needs (Skalla et al. 2004; Dikken & Sitzia 1998). Patients want more specific information on potential toxicities of treatment, such as what side-effects they are likely to experience, their severity, and duration and how to manage them (Skalla et al. 2004). The provision of such information is likely to make them feel more in control of their disease by knowing what to expect and how to deal with problems when they occur. Furthermore, it may prevent unnecessary worry and anxiety over side-effects that are less likely to arise (Skalla et al. 2004). Whilst not only having the potential to greatly improve the experiences of patients with cancer receiving chemotherapy, a more accurate prediction of potential side effects of such treatments, based on patients individual and disease related characteristics, may also be of significant benefit clinically. By knowing the likelihood of potential side effects occurring, patients who are 'at risk' of developing certain toxicities may be identified, facilitating more targeted and cost effective interventions, to those in greatest need and who are most likely to benefit. It may also guide clinicians in the selection of appropriate treatments for individual patients based on their characteristics and needs.

2.2 Risk Modelling

Within health care, there is increasing use of predictive models to identify patients who are most likely to experience specific disease and/or treatment related events. Relative to cancer care, such models have tended to focus on predictors of survival and life threatening toxicities such a febrile neutropenia (Chow, Harris, & Fung 2006; Donohue 2006; Lyman et al. 2005; Sanchez et al. 2006; Vigano et al. 2000). In relation to the prediction of symptoms, there is limited work that has been performed in this area, particularly in relation to the area of chemotherapy side-effects (Armer et al. 2003; Talcott et al. 2003; Poleshuck et al. 2006).

Risk modelling provides a powerful mechanism for identifying patterns in data and predicting what will happen in the future. A variety of techniques can be employed to analyse data and the results of such analysis can be used to provide likelihood information relating to the prevalence of similar data occurring in the future. This information can relate to the likelihood of specific data values occurring together, or perhaps the frequency with which whole records of information may occur again.

The potential for using mathematical techniques to identify risk is reinforced by their prevalence in the literature: Cowie et al (2006) discussed the use of Bayesian belief networks in aiding in dementia diagnosis, where patterns are identified in data from patients who are potentially dementia sufferers and such patterns used to help predict whether dementia is present; Werner and Fogarty (2001) developed mathematical models to allow simulation of future events based on past medical records. Using this technique, the occurrence of thrombosis was predicted in sufferers of collagen disease; De Toro et al (2003) used neural networks to predict hospital

mortality of patients in intensive care more accurately than traditional regression models; Dybowski et al (1996) successfully applied multi-objective optimisation to analyse electrocardiogram (ECG) traces to provide a non-invasive technique for diagnosing potential signs of atrial disease; García-Pérez et al (1998) use data mining and neural network techniques and Mani et al (1997) apply decision-trees and rule-based approaches to differentiate between different dementias types.

The risk modelling techniques available differ in the way in which the data is analysed, how much data the technique requires to make significant predictions, and how much information is revealed regarding the patterns that exist. In general, it is advisable to use a variety of different methods to ensure that as much prediction data can be obtained as possible.

2.3 Project Aims

In order to identify patterns in side-effects experienced by patients receiving chemotherapy, data was analysed with a view to answering the following key questions:

- Does the chemotherapy regime impact on the side-effects experienced?
- Can we predict later symptoms from the pattern of early symptoms?
- If side-effects are experienced in an early cycle does this increase the likelihood of experiencing the same side-effects in later cycles?
- Do some side-effects always occur together and does the presence of some symptoms make others less likely to occur?
- Does the severity to which a side-effect is experienced impact on the likelihood of that side-effect occurring again?

The principal aim of the project was to provide new patients with a prediction of side-effects they are likely to experience across all cycles of chemotherapy. The secondary aim was to provide patients with ongoing side-effect information. By monitoring their side-effects over a period of time, we hoped to provide up-to-date predictive information which is revised and reviewed (according to how their side-effects change) over the cycle of treatments. Currently, the study focuses on six symptoms associated with chemotherapy: mucositis, nausea, vomiting, fatigue, diarrhoea, and hand-foot syndrome.

2.4 Research Methodology

2.4.1 Data Collection

Thirty-three retrospective cases of patients with breast cancer undergoing chemotherapy have been used in the study. Risk modelling analysis was performed on this data in an attempt to answer the questions posed in Section 2.2. Current data collection is also taking place from three sites across Scotland, which will form a prospective data set consisting of forty patients. These patients have been diagnosed with breast cancer and are commencing adjuvant chemotherapy.

Data is being collected using a series of daily patient self-reporting paper-based symptom questionnaires collected throughout 4 cycles of chemotherapy. The daily symptom questionnaire is being used in addition to the clinical use of two existing questionnaires commonly used in practice to assess and grade chemotherapy related symptoms – the Common Toxicity (CTC) grading system (National Cancer Institute, 2003) and the Chemotherapy Symptom Assessment Scale (C-SAS) (Brown et al, 2001). This data will be used to further assess the accuracy of the risk-modelling tool.

2.4.2 Data Analysis

The data analysis performed uses both traditional statistical techniques and a class of more advanced, powerful techniques collectively known as 'data mining'. Data mining is task oriented, which means that analysis begins with the definition of a task and progresses through the use of data and software to develop a system for performing the chosen task. In this study, the task is to predict future symptoms from a combination of patient data and current symptoms. The pattern of symptoms experienced as a patient progresses through a chemotherapy regime is not random, and as such can be predicted. Data mining tools are designed to find the structure that allows such predictions to be made.

The data from this study was analysed in two distinct forms. One which treats the data as a time series on the assumption that there is a pattern in the way symptoms evolve over time (trends or cycles, for example) and the other being static, working on the assumption that the patient's initial state and the chemotherapy regime alone are sufficient to predict when and with what severity symptoms will occur. For the time series analysis, we used dynamic Bayesian networks, Markov models and a decompositional approach. For the static prediction the principal tools used were neural networks, cluster analysis, and Bayesian belief networks.

3 ASYMS©-SERAT

3.1 Introduction to ASyMS©-SERAT

The ASyMS©-SERAT tool will be incorporated into a mobile phone based, advanced symptom management system (ASyMS©-C) which has been developed to remotely monitor the side effects of chemotherapy in patients with cancer receiving chemotherapy (Maguire et al, 2005).

The ASyMS©-SERAT tool employs the use of risk modelling techniques to provide patients and clinicians with predictions of likely side effects. The prototype tool can be used to provide predictive information to both new patients, and those currently undergoing treatment. For new patients, the tool allows patient specific data to be entered and provides feedback as to likely side-effects (along with severity details) that will occur. As patients undergo treatment and side effects are monitored, the tool can measure these against the original prediction model. Prior to each cycle of treatment, a patient's predicted model can be reviewed and revised to provide a new predictive model if felt necessary.

The development of the tool has been split into two phases: Phase I which concentrates on the provision of information for new patients, and Phase II which provides information for returning patients part-way through their chemotherapy regime. To date, Phase I has been completed and it is envisaged that Phase II will be completed by February 2008.

3.2 ASyMS©-SERAT in Use

This description of ASyMS©-SERAT will focus on Phase I of the tool as this has now been completed. In Phase I of ASyMS©-SERAT, the tool uses information it has learnt from the data and combines this with patient specific information to predict the likely side effects a patient will experience over the course of their treatment. The patient can receive predictions relating to possible symptoms they are likely to experience in their first cycle of treatment as well as possible symptoms they are likely to experience across all cycles of treatment.

In Figure 1 we provide a sample screenshot depicting information about side-effects a patient may experience in their first cycle of treatment. From the textual scrolling area at the top of the screen, it is evident that this patient has a high chance of experiencing both nausea and fatigue during cycle 1, and some possibility of experiencing mucositis. The pie charts show how severe each of these symptoms could be. For example, about 17% of cases of nausea will be severe, about 33% will be moderate, and about 50% will be mild.

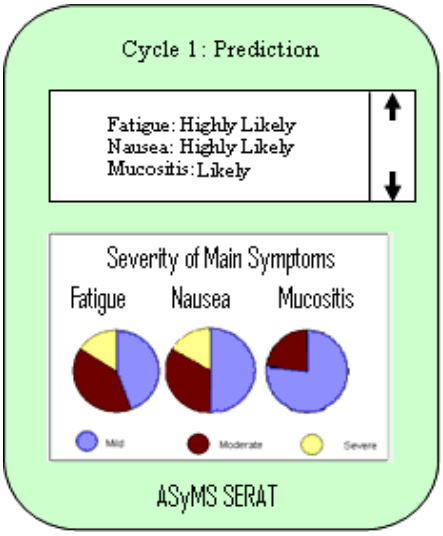


Figure 1: Screenshot of ASyMS©-SERAT showing likely side-effects in cycle 1 of treatment

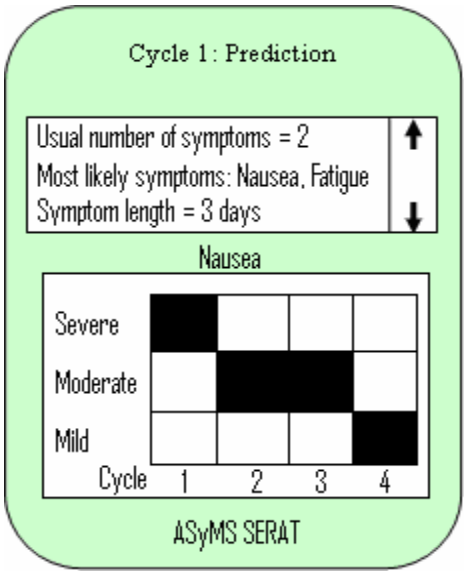


Figure 2: Screenshot of ASyMS©-SERAT showing predicted severity of nausea symptoms across cycles 1-4.

The screenshot in Figure 2 provides longitudinal information across all cycles. Such information can be used to inform the patient about likely patterns of symptoms over time. In the example shown, it is evident that although the patient is experiencing severe nausea in cycle one, this symptom will become moderate during cycles two and three, falling to mild by cycle four.

4 CONCLUSIONS AND FUTURE WORK

Although research on the project is still in its infancy and the ASyMS©-SERAT tool is very much a prototype system, initial results from the risk modelling analysis are very promising. From initial testing it would seem that through use of ASyMS©-SERAT, accurate, personalised predictions of possible side-effects can be made, providing patients with a more informed view of their treatment, and clinicians with the information required for preventative measures or management of side-effects to be applied where possible.

Once Phase II of ASyMS©-SERAT tool is complete, we hope to incorporate the tool in existing ASyMS© symptom management software. This complete symptom prediction and management tool will hopefully allow patients to feel more in control of their symptoms, knowing in advance what to expect, and how to manage the symptoms accordingly. A larger, more comprehensive evaluation of the ASyMS©-SERAT tool will be conducted as part of this work. We are currently in the process of applying for further funding to facilitate this next stage of the project.

ACKNOWLEDGEMENTS

We are grateful to the following organisations for funding this project: Stirling University Research and Enterprise (SURE) Ltd. and Fife & Forth Valley Enterprise.

REFERENCES

Armer, J. M., Radina, M. E., Porock, D., & Culbertson, S. D. 2003, "Predicting breast cancer-related lymphedema using self-reported symptoms", *Nursing Research*, vol. 52, no. 6, pp. 370-379.

Chen-Hardee, S., Chrischilles, E. A., Voelker, M. D., Brooks, J. M., Scott, S., Link, B. K., & Delgado, D. 2006, "Population-based assessment of hospitalizations for neutropenia from chemotherapy in older adults with non-Hodgkin's lymphoma (United States)", *Cancer Causes & Control*, vol.17, no.5, pp.647-654.

Chow, E., Harris, K., & Fung, K. 2006, "Sucessful validation of a survival prediction model in patients with metastases of the spinal column", *International Journal of Radiation Oncology, Biology, Physics*, vol. 65, no. 5, pp. 1522-1527.

Cohn, K. H. 1982, "Chemotherapy from the insiders perspective.", *The Lancet*, vol. 319, pp. 1006-1009.

Cowie J., Oteniya, L., Coles, R. 2006 DIAGNOSIS OF DEMENTIA AND ITS PATHOLOGIES USING BAYESIAN BELIEF NETWORKS. In Proceedings of the 8th International Conference on Enterprise Information Systems: Artificial Intelligence and Decision Support Systems, INSTICC Press, Paphos, Cyprus. pp 291-295.

De Toro F, Ros E, Mota S, Ortega J. 2003 Non-invasive Atrial Disease Diagnosis Using Decision Rules: A Multi-objective Optimisation Approach. In: Fonseca CM, Flemming PJ, Zitzler E, Kalyanmoy D, Thiele L, eds. Evolutionary Multi-Criterion Optimization. Proc. of 2nd Int. EMO Conference, Faro, Portugal. Springer: 638-647.

Dikken, C. & Sitzia, J. 1998, "Patients' experiences of chemotherapy: side-effects associated with 5-flurouracil and folinic acid in the treatment of colorectal cancer", *Journal of Clinical Nursing*, vol. 7, pp. 371-379.

Donohue, R. B. 2006, "Development and Implementation of a Risk Assessment Tool for Chemotherapy-Induced Neutropenia", *Oncology Nursing Forum*, vol. 33, no. 2, pp. 347-352.

Du, X. L., Osborne, C., & Goodwin, J. S. 2002, "Population-based assessment of hospitalizations for toxicity from chemotherapy in older women with breast cancer", *Journal of Clinical Oncology*, vol. 20, no. 24, pp. 4636-4642.

Dybowski R, Weller P, Chang R, Gant V. Prediction of outcome in the critically ill using an artificial neural network synthesised by a genetic algorithm. The Lancet (vol 347): Elsevier, 1996: 1146-1150.

García-Pérez E, Violante A, Cervantes-Pérez F. 1998 Using neural networks for differential diagnosis of Alzheimer disease and vascular dementia. Expert Systems with Applications (vol.14). Elsevier; 219-225.

Kuderer, N. M., Dale, D. C., Crawford, J., Cosler, L. E., & Lyman, G. H. 2006, "Mortality, morbidity, and cost associated with febrile neutropenia in adult cancer patients", *Cancer*, vol. 106, no. 10, pp. 2258-2266.

Lyman, G. H., Lyman, C. H., Agboola, O., & for the ANC Study Group 2005, "Risk Models for Predicting Chemotherapy-Induced Neutropenia", *The Oncologist*, vol. 10, pp. 427-437.

Maguire, R., Miller, M., Sage, M., Norrie, J., McCann, L., Taylor, L., & Kearney, N. 2005, "Results of a UK based pilot study of a mobile phone based advanced symptom management system (ASyMS) in the remote monitoring of chemotherapy related toxicity", *Clinical Effectiveness in Nursing*, vol. 2005 Sep-Dec; 9, no. 3-4, pp. 202-210

Mani S, Shankle WR, Pazzani MJ, Smyth P, Dick MB. 1997 Differential Diagnosis of Dementia: A Knowledge Discovery and Data Mining (KDD) Approach. In: Masys DR, ed. Journal of American Medical Informatics Association supplement. Full paper in extended proceedings of Proc. of AMIA., Hanley and Belfus : 875-880.

McCaughan, E. M. & Thompson, K. A. 2000, "Information needs of cancer patients receiving chemotherapy at a day-case unit in Northern Ireland",

Journal of Clinical Nursing, vol. 2000 Nov; 9, no.6, pp. 851-858.

National Cancer Institute (2003) Common Toxicity Adverse Events Criteria

Polednak, A. P. 2004, "Surveillance for hospitalizations with infection-related diagnoses after chemotherapy among breast cancer patients diagnosed before age 65", *Chemotherapy.Vol.50(4)()(pp 157-161), 2004.* no. 4, pp. 157-161.

Poleshuck, E. L., Katz, J., Andrus, C. H., Hogan, L. A., Jung, B. F., Kulick, D. I., & Dworkin, R. H. 2006, "Risk factors for chronic pain following breast cancer surgery: a prospective study", *Journal of Pain*, vol. 7, no. 9, pp. 626-634.

Sanchez, C. d. M., Elustondo, S. G., Estirado, A., Sanchez, F. V., Cooper, C. G., Romero, A. L., Otero, A., & Olmos, L. G. 2006, "Palliative Performance Status, Heart Rate and Respiratory Rate as Predictive Factors of Survival Time in Terminally Ill Cancer Patients", *Journal of Pain & Symptom Management*, vol. 31, no. 6, pp. 485-492.

Skalla, K. A., Bakitas, M., Furstenberg, C. T., Ahles, T., & Henderson, J. V. 2004, "Patients' need for information about cancer therapy", *Oncology Nursing Forum*, vol. 2004 Mar; 31, no. 2, pp. 313-319.

Talcott, J. A., Manola, J., Clark, J. A., Kaplan, I., Beard, C. J., Mitchell, S. P., Chen, R. C., O'Leary, M. P., Kantoff, P. W., & D'Amico, A. V. 2003, "Time Course and Predictors of Symptoms After Primary Prostate Cancer Therapy", *Journal of Clinical Oncology*, vol. 21, no. 1, pp. 3979-3986.

Tierney, A. J., Taylor, J., & Closs, S. J. 1992, "Knowledge, expectations and experiences of patients receiving chemotherapy for breast cancer.", *Scandanavian Journal of Caring Sciences*, vol. 6, no. 2, pp. 75-80.

Vigano, A., Dorgan, M., Buckingham, J., Bruera, E., & Suarez-Almazor, M. 2000, "Survival prediction in terminal cancer patients: a systematic review of the medical literature", *Palliative Medicine*, vol. 14, pp. 363-374.

Werner J.C & Fogarty T.C. 2001. Genetic programming applied to Collagen disease and thrombosis. In: De Raedt L, Siebes A, eds. Discovery challenge on Thrombosis Data. European Conf. on Principles and Practice of Knowledge Discovery in Databases. Springer-Verlag, 14-20.

STAINING PATTERN CLASSIFICATION IN ANTINUCLEAR AUTOANTIBODIES TESTING

Paolo Soda and Giulio Iannello
Facoltà di Ingegneria, Università Campus Bio-Medico di Roma, Via Alvaro del Portillo 28, Roma, Italy
p.soda@unicampus.it, g.iannello@unicampus.it

Keywords: Computer Aided Diagnosis (CAD, Multiple Expert Systems, Classifier Aggregation, Medical Imaging, Indirect ImmunoFluorescence (IIF), HEp-2 Cell Classification.

Abstract: In Indirect Immunofluorescence (IIF) the use of Computer-Aided Diagnosis (CAD) tools can support physicians' estimation of both fluorescence intensity and staining pattern. This paper reports our experiences in the staining pattern recognition of IIF wells. Since several cells constitute each well, we have developed a Multiple Expert System (MES) based on the one-per-class approach devised to classify the pattern of individual cells. As a novelty, we introduce an aggregation rule based on the estimation of the reliability of each composing experts. Then, the whole well staining pattern is computed using the reliability of its cells classification. The approach has been successfully tested on an annotated set of IIF images.

1 INTRODUCTION

Connective tissue diseases (CTD) are autoimmune disorders characterized by a chronic inflammatory process involving connective tissues. Detection of antinuclear antibodies (ANA) is a common marker in patients with suspected CTD. The recommended method for ANA testing is the Indirect Immunofluorescence (IIF) microscopy based on HEp-2 substrate (Center for Disease Control, 1996). IIF slides are examined at the fluorescence microscope, and physicians report both the fluorescence intensity classification and the staining pattern description. The former is scored semi-quantitatively with respect to both positive and negative controls contained in each slide (Center for Disease Control, 1996). The latter is reported only for positive samples, since they may reveal different patterns of immunofluorescent staining that are relevant to diagnostic purposes. Indeed, more than thirty different nuclear and cytoplasmic patterns could be identified, which are given by upwards of one hundred different autoantibodies. In the literature such patterns are typically grouped in the following classes (Rigon et al., 2007; Sack et al., 2003), that are specific to the most relevant and recurrent autoantibodies: (a) *Homogeneous*, staining of the interphase nuclei and of the mitotic cells chromatin; (b) *Peripheral nuclear* or *Rim*, staining around the outer region of the nucleus, with weaker staining toward the center; (c) *Speckled*, fine or coarse granular nuclear staining of the interphase cell nuclei; (d) *Nucleolar*, large coarse speckled staining within the nucleus, less than six in number per cell; (e) *No pattern*: unclassifiable pattern. Figure 1 depicts four examples of easily distinguishable staining patterns. No instance of the *no pattern* class is reported since it is quite impossible to find positive wells belonging to that class, whereas some cells without a classifiable pattern can occur in a given well.

The staining patterns may be evaluated at various dilutions. On the one hand at high titer, e.g. 1:160, they are usually clearly describable even if they contemporaneously occur, since only very positive sera exhibit detectable fluorescence intensity. On the other hand low dilutions, e.g. 1:40, allow detecting weak positive sera. An intermediate 1:80 dilution is the recommended and the most used one (Center for Disease Control, 1996), even because it allows not carrying out the end-point dilution[1]. At such a titer the staining patterns are not easily detectable since both strong and weak positive sera are positive. Indeed, for the former sera the staining pattern is usually evident, whereas for the latter it is not noticeable.

In the field of autoimmune diseases, the availability of accurately performed and correctly reported laboratory determinations is crucial for the clinicians, demanding for highly specialized personnel that are

[1]End-point dilution consists of patient serum progressive dilution, until the fluorescence intensity disappears. It is very expensive in time and cost, because the analysis of a single patient requires more than a well.

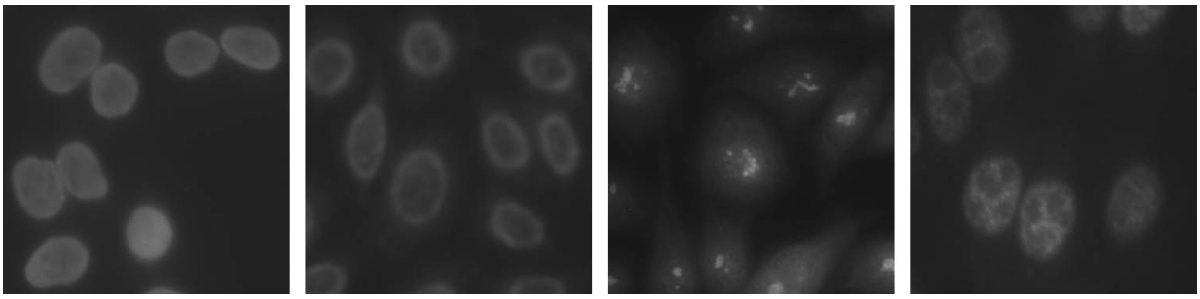


Figure 1: Examples of homogenous, peripheral nuclear, nucleolar and speckled images (left to right).

not always available. Moreover, the readings in IIF are subjected to interobserver variability that limits the reproducibility of the method (Rigon et al., 2007; Sack et al., 2003). To date, the highest level of automation in IIF tests is the preparation of slides with robotic devices performing dilution, dispensation and washing operations (Das s.r.l., 2004; Bio-Rad Laboratories Inc., 2004).

In this respect, some papers recently proposed Computer-Aided Diagnosis (CAD) systems that support physicians' decision on both the two sides of IIF test diagnosis, i.e. the fluorescence intensity (Soda and Iannello, 2006) and the staining pattern (Sack et al., 2003; Perner et al., 2002; Soda, 2007).

With reference to the latter recognition task, in (Sack et al., 2003) and (Perner et al., 2002) the authors presented a decision tree induction algorithm based on texture features extracted from segmented cells of wells diluted at 1:160. These systems, which classify the single cells pattern, exhibited an error rate (P_e) of 16.9% (Sack et al., 2003) and 25.6% (Perner et al., 2002). Recently, in (Soda, 2007) we presented a recognition system devised to label the pattern of HEp-2 wells diluted at 1:80 on the strength of its cells classification. To recognize the individual cells, we employed a classifier aggregation approach that introduced a fixed reject rate since the decision rule was very conservative (P_e and the reject rate (P_r) were 13.1% and 42.6% for individual cells, respectively). To determine the staining pattern of the whole well, two different procedures based on relative and absolute majority of cells distribution among the various classes were proposed. Using the former rule, $P_e = 12.9\%$ and $P_r = 16.6\%$, whereas adopting the latter rule $P_e = 8.3\%$ and $P_r = 29.5\%$, respectively.

Having developed a CAD for fluorescence intensity, and motivated by the need to overcome the current IIF limitations, here we present a system that supports the classification of the staining pattern of the whole well at the 1:80 titer. With respect to (Soda, 2007), we use a more sophisticated classification architecture, a different selection rule and criterion to determine the well pattern in order to improve the recognition performance. Furthermore, our work differs from (Sack et al., 2003) and (Perner et al., 2002) for three main reasons: first, such papers classified only the pattern of individual cells and not that of the whole well; second, they used samples diluted at 1:160 and, third, our data set contains only positive wells. With reference to the last point, notice that both the medical guidelines and the clinical practice require labelling the staining pattern only for positive samples.

2 OUR APPROACH

To classify the well staining pattern into one of the basic groups reported in the previous section, first we segment the image to locate the cells and extract the set of features; second, we classify the staining pattern of cells and, third, we classify the staining pattern of the whole well on the strength of the classification of its cells.

In our opinion, such an approach addresses some key-points of IIF staining pattern classification. Indeed, a recognition approach based on the classification of individual cells has the potential for detecting the occurrence of multiple patterns, i.e. the predominant and the minor ones. Furthermore, this approach is tolerant with respect to misclassifications in cells recognition, since the final label of the well is computed by using several pieces of information. Indeed, if enough cells per well are available, it is reasonable that cells misclassification, if limited, does not affect the well pattern classification.

On the basis of a system that classifies the pattern of individual cells, different rules can be employed to label the well staining pattern. Differently from (Soda, 2007), here we apply the Weight Sum Rule (WS) among the classification of each cell, which is

given by:

$$WS_i = \sum_X \phi(x) \cdot I_i(x) \quad (1)$$

where $\phi(x)$ indicates the classification reliability of input cell x and $I_i(x)$ denotes an indicator variable defined as follows:

$$I_i(x) = \begin{cases} 1 & \text{if the cell } x \text{ belongs to class } C_i \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

The index of the final class of well staining pattern is $\upsilon = \arg\max_i(WS_i)$, i.e. the class for which WS_i is maximum.

3 ARCHITECTURE OF CELLS RECOGNITION SYSTEM

Preliminary results on the classification of the staining pattern of individual cells suggested us to use a combination of experts rather than a single one. In this respect, in the literature it has been observed that the recognition performance attainable combing set of classifiers, as well as different features, should be improved by taking advantages of the strengths of the single experts, without being affected by their weakness.

As recognition system we employ a Multi-Expert Systems (MES) based on the *one-per-class* paradigm, which assumes that the multiclass learning problem is reduced to several binary classification tasks (Jelonek and Stefanowski, 1998; Allwein et al., 2001). Given the number L of classes in which the input samples are distributed, the MES is composed by L modules, each one being an expert in the recognition of one input class from the other (part *A* of figure 2). The base blocks should be considered complementary rather than competitive. Their predictions are aggregated to a final classification decision on a basis of a given rule. Indeed, in the figure the individual decisions are given to an *aggregation module*, which identifies the block that is the most likely to be correct for any input sample.

The rationale of such an architecture is inspired by the results coming out from the feature selection phase: the set of stable and effective features obtained for each class enforced the evidence that the classification could be reliably faced by introducing one specialized module per each class that the system should recognize.

From a theoretical point of view, each module of *A* in figure 2 can be constituted either by a single classifier or by employing again a multiple experts scheme. In the latter case, the classifiers combination can be based on both fusion (e.g. as in (Kuncheva et al., 2001; Gunes et al., 2003)) and selection techniques (e.g. as in (Xu et al., 1992; Giacinto and Roli, 2001)), or on a mixture of them (e.g. as in (Kuncheva, 2002)). The fusion scheme supposes that all classifiers are equally "skilled" and applied in parallel over the whole feature space, providing robustness by multiplying the number of observation channels, which are then combined in a data fusion block. The selection scheme, assuming that each classifier is an expert in some local area of the feature space, identifies which expert has the biggest accuracy in a local region surrounding the sample, letting it label the input. To improve the recognition performance attainable by the L modules, we implement them with multiple binary classifiers combined by fusion, as depicted in part *B* of figure 2. Specifically, as a fusion rule we use Weighted Voting (WV) (Cordella et al., 1999).

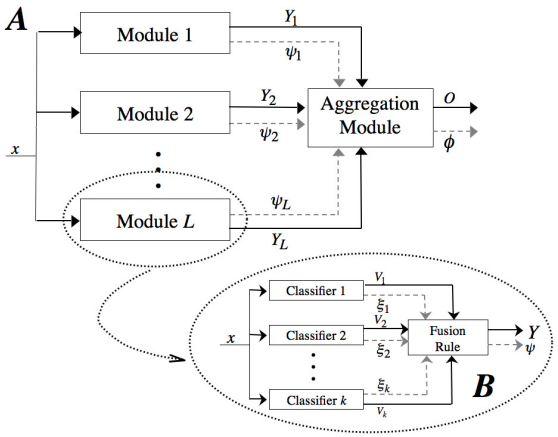


Figure 2: The system architecture. Part ***A***: aggregation of binary modules. Part ***B***: each module is composed by a fusion of experts.

The overall resulting system architecture combining the different MES schemes will be referred to as Hybrid-Classifier-Aggregation-Fusion (HCAF). In the following we adopt a top-down approach to further present our recognition system: first we report the rule applied in the aggregation module and then we describe the fusion strategy internal to each block.

Classifier Aggregation. The rule evaluates which single module is most likely to be correct for any given sample. Since each module has a binary output, possible input combinations to the aggregation module can be grouped into three categories: (i) those for which only one module j classifies the sample in its class C_j, (ii) those for which more modules classify the sample in its own class, (iii) those for which none module classifies the sample in its class.

We introduce a strategy based on reliability estimation that chooses an output, $O(x)$, in any of the possible combinations of modules' output, referred to as Reliability-based-Aggregation (RbA). The rationale lies in the observation that such an evaluation is useful for solving complex pattern recognition tasks (Cordella et al., 1999). Let us then denote $\psi_j(x)$ and $Y_j(x)$ the reliability parameter and the output of the jth module when it classifies the sample x, respectively. Since in case (i) all the modules agree in their decision, as a final output is chosen the class of the module whose output is 1. Conversely, in cases (ii) and (iii) the final decision is performed looking at the reliability of each modules' classifications.

More specifically, in case (ii), m modules vote for their own class, with $2 < m \leq L$, whereas the others $(L-m)$ ones indicate that x does not belong to their own class (i.e. their outputs are 1 and 0, respectively). To solve the dichotomy between the m conflicting modules we look at the reliability of their classifications and choose the more reliable one. Formally:

$$O(x) = C_j, \text{where } j = \arg \max_{i:Y_i(x)=1} (\psi_i(x)) \quad (3)$$

In case (iii), all modules classify x as belonging to another class than the one they are specialized in (i.e. their outputs are 0). In this case, the bigger is the reliability parameter $\psi_j(x)$, the less is the probability that x belongs to C_j, and the bigger is the probability that it belongs to the other classes. These observations suggest selecting the following selection rule:

$$O(x) = C_j, \text{where } j = \arg \min_{i:Y_i(x)=0} (\psi_i(x)) \quad (4)$$

In other words, we first find out which module has the minimum reliability and then we choose the class associated to it as a final output.

Classifier Fusion. Each specialised module of the system is composed by an ensemble of classifiers combined by the Rule (WV) (part *B* of figure 2). In such a procedure, each expert gives its opinion, i.e. a vote, about the class of the input pattern, which is then weighted by a reliability parameter. If we denote as $\xi_k(x)$ the value of reliability of kth classifier on sample x and $V_k^i(x)$ the vote for class C_i of kth classifier on sample x, the weighted sum of votes for class is given by:

$$W_i(x) = \sum_k \xi_k(x) \cdot V_k^i(x) \quad (5)$$

Therefore, the output of WV rule, $Y(x)$, is the index of class $C_Y(x)$ for which $W_i(x)$ is maximum:

$$Y(x) = \arg\max_i (W_i(x)) \quad (6)$$

Note that to estimate $\xi_k(x)$, all classifiers have to work at a *measurement* level, i.e. they attribute each class a measurement value representing the degree that the input sample belongs to that class.

4 RELIABILITY ESTIMATORS

The approach previously described requires the introduction of parameters that estimate the classification reliability of both individual expert and fusion of experts as well as the overall cells classifier, named as ξ, ψ and ϕ, respectively. Note that all of them vary in the interval $[0,1]$, and a value near 1 indicates a very reliable classification. The first issue, i.e. the definition of estimators that compute the reliability of each classification act for measurement classifiers, has been discussed in the literature (Cordella et al., 1999; De Stefano et al., 2000). For their formal definition in case k-Nearest-Neighbour (kNN) and Multi-Layer Perceptrons (MLPs), i.e. the single classifiers used in the present work, see (Cordella et al., 1999). Note that such formulas have proven usefulness also in other application, e.g. in (De Stefano et al., 2000).

The reliability ψ of WV classification has been computed according to a method similar to the one reported in (Cordella et al., 2000), which is based on the estimation of maximum reliability for the winning class and for the others classes, respectively. Note that ψ is calculated for each input samples and it is then used in the aggregation module to determine the cell final output $O(x)$.

The reliability estimation for the classification of each input cell performed by the overall MES is required to determine υ, i.e. the index of the well staining pattern class, as presented in section 2. In this respect, the overall reliability ϕ considers not only the reliability ψ of the selected module, but also the reliabilities of the other blocks (Soda and Iannello, 2007). For all the three input combinations to the aggregation module, i.e. (i), (ii) and (iii), such a choice accurately estimate the classification reliability of each sample, since it considers the agreement between all modules. For the sake of brevity, we do not report the details here. The interested reader may find them in (Soda and Iannello, 2007).

5 DATA SET

To populate a referring data set, we use 37 images of positive wells , grouped as follows: 24.3% are *Homogeneous*, 21.6% are *Peripheral nuclear*, 35.1% are *Speckled*, 18.9% are *Nucleolar*.About 15 segmented

cells per well are chosen at random, located as reported in (Soda and Iannello, 2006) and then cropped to a rectangular region.

To develop the MES devised to recognize individual cells, we need their labels that are determined by two specialists at a workstation monitor. To this aim, the classes introduced in section 1 do not cover all the possibilities. Indeed, on the one hand those classes represent a global pattern, i.e. the pattern of whole well that is given by the global observation of several cells. On the other hand, each cell could potentially show a staining pattern that is different from the well pattern. To overcome such limitations, for manual labelling we adopt the following classes, as reported elsewhere (Perner et al., 2002) (for definition of classes (i)–(iv) and (viii) see section 1): (i) *homogeneous* (HO), (ii) *peripheral nuclear* or *rim* (PN), (iii) *speckled* (SP) (iv) *nucleolar* (NU), (v) *artefact* (AR), i.e. cell corrupted during the slide preparation process, identifiable with an irregular shape, (vi) *positive mitosis*, i.e. the nonchromosome region of metaphase mitotic cells demonstrate staining, (vii) *negative mitosis*, i.e. the nonchromosome region of metaphase mitotic cells is negative, (viii) *no pattern* (NP). Since the number of cells belonging from groups (vi)–(viii) in not statistically meaningful, they are not considered in the following.

The data set consists of 573 labelled cells, therefore subdivided: 23.9% HO, 21.8% PN, 37.0% SP, 8.2% NU and 9.1% AR.

To analyze the staining pattern we compute a set of features related to texture components, adopting both statistical and spectral features. The former measures are associated to properties of the first and the second order histogram, respectively. The spectral features are calculated by partitioning the spectrum of the Fourier Transform into angular and radial bins. Furthermore features related to Wavelet Transform and Zernike Moments have been computed. Results of discriminant analysis show that all the extracted features have limited discriminant strength over five classes (i.e. HO, PN, SP, NU and AR), but different feature subsets discriminate better each class from the others, enforcing the rationale of adopting the one-per-class approach.

6 RECOGNITION RESULTS

With reference to the classification of individual cell, the HCAF system is a MES constituted by five modules each one devised to recognized one of the five input classes. i.e. HO, PN, SP, NU and AR. Each block is composed by a fusion of individual classifiers, such as kNN and MLP combined by the WV algorithm.

Table 1: Confusion matrix of HCAF classifier employing the reliability-based selection (RbA) rule.

		Input class				
		HO	PN	SP	NU	AR
Output class	HO	73.9%	5.6%	5.2%	8.5%	15.4%
	PN	10.0%	71.2%	3.8%	14.9%	13.5%
	SP	10.2%	12.8%	88.2%	0.0%	17.3%
	NU	1.5%	2.4%	0.5%	72.3%	9.6%
	AR	4.4%	8.0%	2.4%	4.3%	44.2%

The HCAF system recognition performance has been evaluated according to a eightfold cross validation approach. They are reported as confusion matrix in table 1. The classification accuracy of HO, PN and NU classes ranges from 71% to 74%, whereas the best and worst recognition performance are attained for cells of SP and AR classes, i.e. 88% and 44%, respectively.

In our opinion, on the one hand, misclassifications of HO, PN and SP samples are related to their similarities of staining pattern and texture. Indeed, the discrimination between such classes is a burdensome issue also for well-trained specialists. On the other hand, errors on NU and NP classes are related to the small cardinality of such sets. Moreover, the variability among AR samples is high, since such class contains those cells corrupted during the slide preparation that exhibit irregular shape and texture. Finally, taking notice of the absolute performance, the 75.9% of cells are correctly classified.

In summary, we observe that the overall performance of the presented cells classifier outperforms that reported in (Soda, 2007). Furthermore, a direct comparison of this results with respect to (Perner et al., 2002) and (Sack et al., 2003) is not possible, since their recognition task differs from ours. Indeed, in those papers the authors used a different data set, which is not only constituted by samples diluted at 1:160, but also containing cells that were negative, i.e. they did not exhibit a detectable fluorescence intensity.

With reference to the performance achieved in the recognition of the whole well staining pattern, note that we have to manage data related to individual cell classification. The *a priori knowledge* based on established medical information excludes the AR cell class from the set of whole well pattern ones (see section 1). Therefore, υ, i.e. the index of well pattern class, is computed from cell class indexes $\{$HO, PN, SP, NU$\}$.

For all the wells, we randomly subdivide their cells into two equal partitions, and then each partition is first used as a training set and then as test set. We deem that such a ration is a good balance between the need of keeping the training set representative as most

as possible and having enough test cells per well to classify the staining pattern in accordance to the WS criterion. In the two trials, the overall system misclassified only one out of the 37 wells, attaining an hit rate equal to 97.3% and outperforming the results reported in (Soda, 2007) (see section 1).

7 CONCLUSIONS

In this paper we have presented a system that supports the staining pattern classification of IIF slides, whose results show high accuracy. The approach, which provides a degree of redundancy that lowers the effect of cell misclassifications, is based on the reliability estimation. The latter is unusual among the classifier aggregation strategies.

We are currently engaged in populating a larger database to consider not only the most relevant and recurrent staining patterns, but also the minor ones. Furthermore, we should apply boosting techniques to improve binary recognition performance, especially in the case of nuclear samples. The research goal is a comprehensive CAD supporting all phases of IIF diagnosis, i.e. both fluorescence intensity and staining pattern classification.

ACKNOWLEDGEMENTS

The authors thank A. Afeltra and A. Rigon for their collaboration in IIF images annotation. This work has been funded by DAS s.r.l of Palombara Sabina (`www.dasitaly.com`).

REFERENCES

Allwein, E. L., Schapire, R. E., and Singer, Y. (2001). Reducing multiclass to binary: a unifying approach for margin classifiers. *J. Mach. Learn. Res.*, 1:113–141.

Bio-Rad Laboratories Inc. (2004). PhD System. USA.

Center for Disease Control (1996). Quality assurance for the indirect immunofluorescence test for autoantibodies to nuclear antigen (IF-ANA): approved guideline. *NCCLS I/LA2-A*, 16(11).

Cordella, L., Foggia, P., Sansone, C., Tortorella, F., and Vento, M. (1999). Reliability parameters to improve combination strategies in multi-expert systems. *Pattern Analysis & Applications*, 2(3):205–214.

Cordella, L., Foggia, P., Sansone, C., Tortorella, F., and Vento, M. (2000). A cascaded multiple expert system for verification. In *1st Int. Work. on Multiple Classifier Systems*, pages 330–339. Springer-Verlag.

Das s.r.l. (2004). Service Manual AP16 IF Plus. Italy.

De Stefano, C., Sansone, C., and Vento, M. (2000). To reject or not to reject: that is the question: an answer in case of neural classifiers. *IEEE Transactions on Systems, Man, and Cybernetics–Part C*, 30(1):84–93.

Giacinto, G. and Roli, F. (2001). Dynamic classifier selection based on multiple classifier behaviour. *Pattern Recognition*, 34(9):1879–1881.

Gunes, V., Menard, M., Loonis, P., and Petit-Renaud, S. (2003). Combination, cooperation and selection of classifiers: a state of the art. *Int. Jour. of Pattern Recognition and Artificial Intelligence*, 17(8):1303–1324.

Jelonek, J. and Stefanowski, J. (1998). Experiments on solving multiclass learning problems by n^2 classifier. In *10th Eur. Conf. on Mach. Learn.*, pages 172–177. Springer-Verlag Lecture Notes in Artificial Intelligence.

Kuncheva, L. (2002). Switching between selection and fusion in combining classifiers: an experiment. *IEEE Transactions on Systems, Man and Cybernetics*, 32(2):146–156.

Kuncheva, L., Bezdek, J., and R.P.W.Duin (2001). Decision template for multiple classifier fusion: an experimental comparison. *Pattern Recognition*, 34:299–314.

Perner, P., Perner, H., and Muller, B. (2002). Mining knowledge for HEp-2 cell image classification. *Journal Artificial Intelligence in Medicine*, 26:161–173.

Rigon, A., Soda, P., Zennaro, D., Iannello, G., and Afeltra, A. (2007). Indirect immunofluorescence (IIF) in autoimmune diseases: Assessment of digital images for diagnostic purpose. *Cytometry - In press*.

Sack, U., Knoechner, S., Warschkau, H., and et al. (2003). Computer-assisted classification of HEp-2 immunofluorescence patterns in autoimmune diagnostics. *Autoimmunity Reviews*, 2:298–304.

Soda, P. (2007). Early experiences in the staining pattern classification of HEp-2 slides. In *Computer Based Medical Systems*, pages 219–224. IEEE Computer Society.

Soda, P. and Iannello, G. (2006). A multi-expert system to classify fluorescent intensity in antinuclear autoantibodies testing. In *Computer Based Medical Systems*, pages 219–224. IEEE Computer Society.

Soda, P. and Iannello, G. (2007). Reliability estimation in the classifier aggregation framework. Technical report, Università Campus Bio-Medico di Roma.

Xu, L., Krzyzak, A., and Suen, C. (1992). Method of combining multiple classifiers and their application to handwritten numeral recognition. *IEEE Transactions on Systems, Man and Cybernetics*, 22(3):418–435.

A GENERIC SOLUTION FOR THE CONSTRUCTION OF DIAGNOSTIC EXPERT SYSTEMS BASED ON PRODUCT LINES

Ma Eugenia Cabello Espinosa and Isidro Ramos Salavert
Polytechnic University of Valencia,Camino de Vera s/n, 46022 Valencia, Spain
mcabello@dsic.upv.es, iramos@dsic.upv.es

Keywords: Expert Systems, Medical Diagnosis, Software Architectures, Reusability of Software, Software Product Lines, Variability, Domain Engineering, Domain Application Engineering, Conceptual Models.

Abstract: This paper presents a generic solution for the construction of diagnostic expert systems using aspect-oriented-software architectures and product line techniques. The approach is shown by specifying a case study using CIMs, and automatically generating a PIM. The case study presented is a medical diagnosis system for the detection of infantile infectious diseases. PRISMA models are used as PIMs. We follow the Model Driven Architecture (MDA) initiative of the Object Management Group (OMG) for building domain models (CIMs), which are automatically transformed into PIMs and are then compiled to a .NET executable application (PSM). The Software Product Line techniques have been used to capture the variability of systems of this kind.

1 INTRODUCTION

In the last few years, there has been an increase in interest in expert systems that perform diagnostic tasks. The main objective of systems of this kind is to capture the state of an entity from a series of data (observation variables) and produce a diagnosis. The domain of expert systems for diagnosis includes systems for medical and education diagnoses, among others. Since systems for medical diagnosis have become more relevant, the need of techniques for their development has also become more important. Additionally, expert systems introduce a difference in the decision making process: they store expert knowledge in a Knowledge Base.

In order to capture complex software requirements, it is also necessy to increase the PIM abstraction level. The PRISMA framework (Pérez, 2006) provides expresivity for specifying software architectures with aspects at a design level. It offers properties and advantages in the construction of complex, distributed, evolutionary, and re-usable architectural models that can be used in the domain of expert systems for medical diagnosis.

Furthermore, the development of these complex systems is becoming more elaborate due to a series of factors. These factors are the emergence of new technologies (Internet and intranet), the interconnection of the different systems and platforms, and to integrate Legacy Systems that are still valid. Also, the need to develop custom software for each type of user complicates the specific aspects of systems in different implementation platforms. This state of affairs requires having multiple versions of the same application in order to deal with all of this variability.

In order to cope with this variability problem, Software Product Lines (SPL) (Clements et al., 2002) have emerged in an effort to control and minimize the high costs of the software development process and to reduce the time to market of these new products. This approach is based on having a base design that is shared by all the product family members. Thus, a specific product can benefit from the common parts of the software. The base design can be re-used in different products by adding different features that caracterize them.

We have built BOM (Base-Line Oriented Modeling), wich is a framework that automatically generates diagnostic systems based on software product lines, to achieve the following goals:

- to create new diagnostic systems in different domains,
- to decrease production costs by reusing software packages or assets,
- to generate automatic code to increase the productivity and quality of software and to decrease the time to market,

- to construct systems in a simpler way by using the ontologies of the diagnosis and the application domains. The models will be closer to the problem domain, which will facilitate user interaction,
- to develop platform independent systems from the problem perspective and not the solution perspective, which will provide generality in the development approach and applicability in different domains.

The products of our Diagnostic Software Product Line (DSPL) have been designed using the PRISMA model and include the architecture and operations of a rule-based expert system.

In this paper the development process of our Diagnostic Software Product Line is presented using a case study of infantile infectious diseases.

The structure of the paper is the following: Section 2 introduces the relevant concepts of Expert Systems, Model Driven Development, the PRISMA Model, and Software Product Lines. Section 3 introduces the variability dimensions of our domain. Section 4 provides an in-depth description of our approach. Section 5 presents a brief summary of the case study. Section 6 presents related works. Section 7 presents our conclusions and provides some ideas for future work.

2 FOUNDATIONS

Our work integrates different technological spaces in order to cope with the complexity of the problem. These are the following:

2.1 Expert Systems

Expert systems capture the knowledge of experts and try to imitate their reasoning processes when the experts solve problems in a certain domain.

These systems usually have a basic architecture that includes a knowledge base, an inference motor, a working memory or facts base, and the user interface. These are the four main components of the architecture of a rule-based expert system.

These components are independent and are composed of separates units. The data are grouped into the working memory (temporary storage of dynamic information). The representation of the knowledge is captured by means of rules of the type: IF <antecedent > THEN <conclusion>, which make up the knowledge base. The control aspect is independent and is performed by the inference motor during the inference processes using different reasoning strategies. The input and output of the information of the systems are done through the user interface.

2.2 Model Driven Development

The Model Driven Development approach (MDD) for software system development is based on the separation of the functionality of the system from its implementation on specific software platforms. MDD increases the abstraction level of the software production process by emphasizing the importance of the conceptual models - Computer Independent Models (CIM) or Platform Independent Models (PIM) - and their role in the software development process.

There are currently two main initiatives in MDD. One of them, which is promoted by the Object Management Group (OMG) is called Model Driven Architecture (MDA) (http://www.org.mda). The other one, promoted by Microsoft, is called Software Factories (SF) and Domain Specific Languages (DSL) (Greenfield et al., 2004).

The key idea of MDA is to focus on the models as first class citizens in the software development process. MDA proposes defining and using models at different abstraction levels. These models can automatically generate code by means of mappings or by applying transformation rules to executable Platform Specific Models (PSM).

2.3 The PRISMA Model

The PRISMA architectural model integrates two approaches: Component-Based Software Development (CBSD) (Szyperski, 1998), and Aspect-Oriented Software Development (AOSD). (http://aosd.net) This integration is obtained by defining the architectural elements through their aspects.

The PRISMA model consists of three types of architectural elements: components, connectors, and systems. A component captures the functionality of the system, whereas a connector acts as a coordinator among other architectural elements. A system is an architectural element of great granularity that allows the encapsulation of a set of components, connectors, and other systems. This, in turn, allows the system to correctly connect them.

PRISMA defines the architectures in two abstraction levels: the type level and the configuration level. In the type level, the types of the architectural artifacts are defined (all of which can be reused): interfaces, aspects, components,

connectors, and systems. In the configuration level, the types are instantiated and the topology of the model (its configuration) is specified.

2.4 Software Product Lines

The Software Product Line (SPL) approach, from a practical point of view, is one of the most successful since it combines systematic development and the reuse of reusable components or assets; i.e., the products are different in some features but share a basic architecture. SPL provides an industrial approach to the software development process.

In the SPL approach, rather than a single application, the development process produce a series or family of them. This implies changing the existing engineering process by introducing a distinction between the domain engineering process and application engineering process. In general, the domain engineering process defines the shared architecture and the variability of the SPL. More than creating products, it is a question of putting assets together in a Base-Line warehouse. For each SPL there is a well defined production plan that specifies the process to obtain each of the individual products. The construction of the assets and their variability (domain engineering process) is separate from the application production (application engineering process).

One of the most important points (or perhaps the most important) of a SPL is the definition of the basic architecture of the Products Line (also called domain architecture). This is due to the fact that this architecture determines the scope of the SPL and the features of the products that can be developed.

One of the key elements for a SPL is how to represent and manage variability. In the context of BOM variability appears in the construction of the domain model (which is represented as a decision tree with different variation points). The base assets are selected by the decision tree. These assets are enriched by the specific features (given in the application model) by a process that results in PRISMA architectural model types. In BOM, the assets are automatically transformed by inserting the instances of the Feature Model (Batory et al., 2006), which gives one executable application.

2.5 Feature Oriented Programming

Feature Oriented Programming (FOP) is the study of the modularity of the features of a domain and their use. The features are considered as first-class citizens in the design process. FOP is an approach to SPL where the programs are built by means of feature composition. These features are considered as building blocks of the programs. They are units that increase monotonically the functionally of the application by providing different products. Each one of the features can be included in the different software artifacts. In general, a SPL is characterized by the set of features being used, which is called the feature model of the SPL.

3 THE VARIABILITY

3.1 Variability in the Diagnostic Domain

After a field analysis of different systems in the diagnostic domain, we can conclude that a diagnosis consists of an interpretation of the involved entity states (as a set of properties), which is followed by the identification of the problem or disfunction by means of its properties. We have detected seven features (or variability sources) that are present in these systems. They are the following:

- a) property views: an entity can have the same properties (the same view), or have different properties (different views) during the diagnostic process,
- b) property levels: the properties of the entities can have *n* different abstraction levels. The rules that relate the proprieties of the entities have *n-1* levels, where n is the level of the proprieties of the entities.
- c) number of hypotheses: the goal of the diagnosis is a single validated hypothesis. There can be one or several candidate diagnostic hypotheses which must all be validated in order to select the appropriate one.
- d) reasoning types: reasoning shows the way in which the rules are applied by the inference motor in order to infer a final diagnosis. The reasoning types can be: deductive reasoning (driven by data), inductive reasoning (driven by goals), and differential reasoning (establishing the difference between two or more diagnostical possibilities),
- e) use case numbers: a use case indicates the division of the system based on its functionality; i.e., the different operations of the systems and how the system interacts with the environment (final users),
- f) number of actors: represents the number of final users of the system,

- g) use cases per actor: a final user can access different use cases.

The variability shown in the use cases, actors, and use cases per actor is reflected in the construction of the architectural elements assets (skeleton) and in the final PRISMA architecture in the following way:

- there is one connector that connects all the architectural component assets for each use case,
- the number of ports of the Inference Motor component is the number of use cases,
- the number of ports of the Knowledge Base is the number of use cases,
- the number of User Interfaces is the number of actors of the use cases,
- the number of ports of the User Interface is the number of use cases that can be accessed by an actor.

The variability of the diagnostic domain is expressed by means of an explicit Features Model.

3.2 Variability in the Application Domain

The features selected from the Features Model must be defined following the case study in which the application is developed. These features are:

- a) name and type of the properties by abstraction level,
- b) rules by abstraction levels: the rules relate the entity properties (name and type) with the properties that are used in the head part and the body part of each one of these rules,
- c) name and type of the hypothesis (and the pre-hypothesis) that are used in the diagnostic process.

4 THE BOM APPROACH

Our work is based in the Software Product Lines approach, on two OMG standards: the Reusable Asset Specification (RAS), which identifies, describes and packs assets in an standard way; and the Software Process Engineering Metamodel (SPEM) (http://www.omg.org/docs/ad/06-06-02.pdf), which defines the standard language for modeling the software process.

In BOM a clear separation is made between the domain engineering and application engineering. This partition is the basis for the reuse and the automation of the software process (Czarnecki et al, 2000). In domain engineering phase, a set of assets and processes are created. In the application engineering phase, these assets are used to produce software products of high quality with a minimal cost and time by executing the stored processes.

4.1 1st Phase: Domain Engineering

In this first phase, the following software artifacts are created by the diagnostic domain engineer. These artifacts are necessary to generate the product plan of our software product line: the artifacts a and b are Computation Independent Models (CIMs), and the artifacts c, d, e, f, g, h, i, j, and k are Platform Independent Models (PIM).

- **a) The (CIM) Features Model**- the Features Model identifies the DSPL in terms of the variability in the domain.

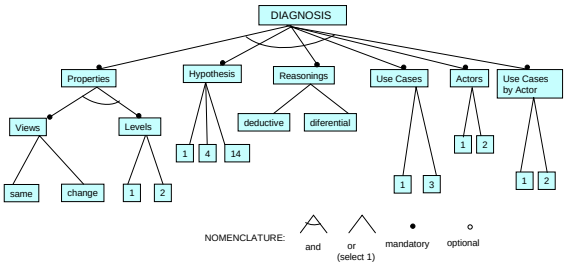


Figure 1: The diagnosis features model.

- **b) The (CIM) Decision Tree**- the sources of variability observed in the Features Model are shown in the variation points of the Diagnostic Decision Tree.

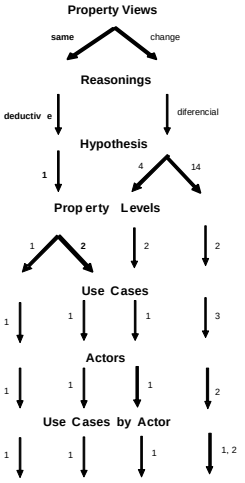


Figure 2: The diagnostic decision tree.

- **c) The (PIM) Domain Conceptual Model**- this model captures the variability of the diagnostic domain.

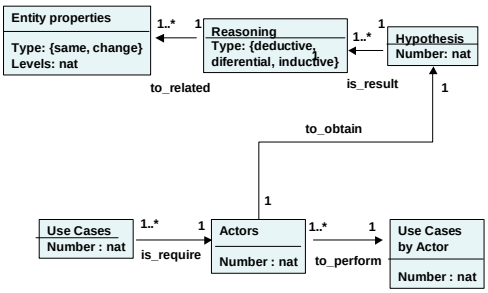


Figure 3: The domain conceptual model.

- **d) The (PIM) Application Domain Conceptual Model**- this model captures the variability of the application domain. (Figure 4 shows the case study of the medical diagnosis).

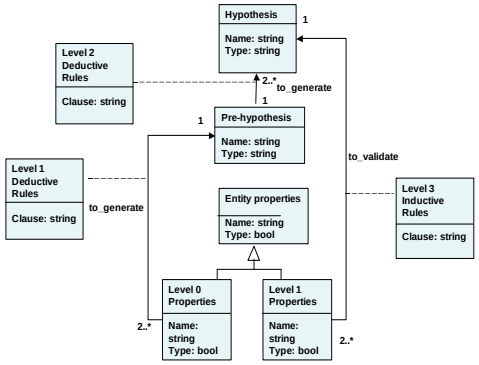


Figure 4: An application domain conceptual model.

- **e) The (PIM) Skeletons or Template Assets**- there are different classes of skeletons or templates for: components, connectors, aspects, and interfaces; they follow the PRISMA Model. The aspects that are necessary for the definition of these architectural elements are: the functional aspects of each one of the components, and the coordination aspect of the connector. These aspects use interface services. These architectural elements are:
  - **The Inference Motor Component**- it establishes the system's control and makes decisions. In addition, the Inference Motor Component provides the general resolution strategy to obtain the diagnosis. It is independent of the system knowledge. This component has a functional aspect that defines the inference process of the system.
  - **The Knowledge Base Component**- it contains the domain knowledge of the case study in rules of inference (Horn clauses) and facts (information that remains unchanged). This component is a temporary warehouse of dynamic information, since the number of facts can be increased as they relate to the inference rules of the domain. When a diagnosis process has concluded, the contents of the work memory is cleared so that the memory is empty before initiating a new diagnosis. The Knowledge Base Component has a functional aspect that defines the domain rules.
  - **The User Component**- it establishes man-machine interaction by allowing communication between the users and the system. Through it, the user offers initial data to the system or answers questions formulated by it. This component has a functional aspect.
  - **The Diagnostic Connector**- this connector synchronizes or requests component's services that are sent/received through its ports. It has a coordination aspect. This diagnostic connector choreographs the diagnostic process.
- **f) The (PIM) Features Insertion Process**- this process inserts the features into the software artefacts. These artefacts represented as XML documents are transformed using XSLT templates.
- **g) The (PIM) RAS Models of the Assets**- its goal is to store information from each one of the assets: ID asset identifier, asset classification, description of the different asset artifacts, variability points of the asset artifacts.
- **h) The (PIM) Assets**- an asset is composed by a skeleton, its RAS Model, and its corresponding insertion process (to be executed in the application construction phase).
- **i) The (PIM) Base-Line**- the Base-Line is the repository that contains all the assets and all the application domain conceptual models, which are used to capture the specific application features.
- **j) The (PIM) Process for selecting the assets and the application domain conceptual models**- This process computes the paths of the decision tree, which is used to select the software artifacts.

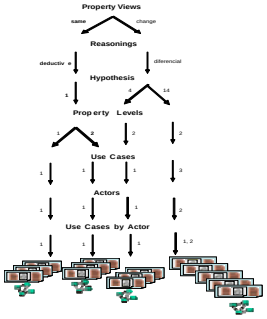


Figure 5: Process for selecting the assets and the application domain conceptual models.

- **k) The (PIM) Process of the production plan of our DSPL**- this process is described using the SPEM.

4.2 2nd Phase: Application Engineering

The production plan of our Software Product Line is described using SPEM. The SPEM Metamodel allows several aspects and problems of the development process to be modeled. In this work, we focus on modeling the tasks, using the SPEM sequence relations without priority. The tasks performed by the application engineer consume input artifacts and produce output artifacts. A task can have associated elements that guide and help in the task execution.

Each one of these tasks is described below:

- **Task 1: To create a configuration of the domain features**- The engineer introduces the domain information of the case study. BOM captures the variability information by means of the domain features detailed in the Domain Conceptual Model. This model allows the engineer (by means of a GUI) to introduce the information of the Model by selecting the product's features using check boxes and pull-down menus.

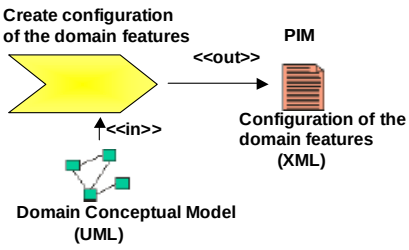


Figure 6: Create a configuration of the domain features.

- **Task 2: To select Assets and the Application Domain Conceptual Model**.- BOM selects the assets and the Application Domain Conceptual Model from the Base-Line (by means of the decision tree).

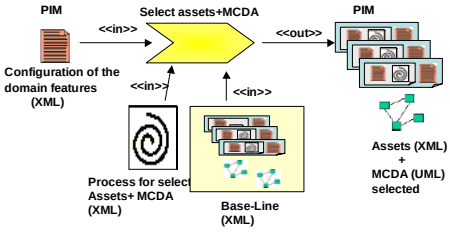


Figure 7: Select assets and application domain conceptual model.

- **Task 3: To create a configuration of the application domain features**.- The engineer introduces the application domain information of the case study. BOM captures the variability information by means of the application domain features contained in the Application Domain Conceptual Model. This model allows the engineer (by means of a GUI) to introduce the information present in that Model using check boxes and pull-down menus.

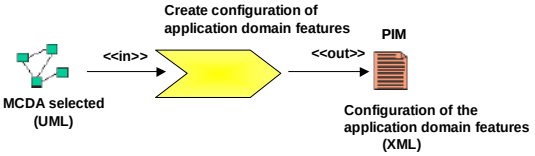


Figure 8: Create a configuration of the application domain features.

- **Task 4: To create PRISMA software artifact types**.- BOM fills the selected skeletons with the data of the specific features of the case study that were defined by the engineer, thereby creating the PRISMA software artifact types. The transformation is represented as XSLT documents to apply on the XML pages representing the software artifacts.

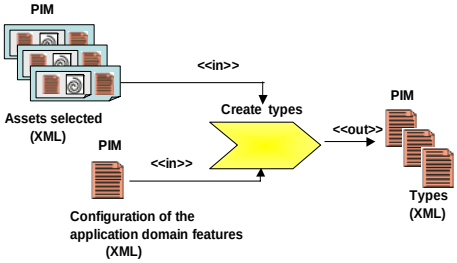


Figure 9: Create PRISMA software artifacts types.

- **Task 5: To configure the Architectural Model-** BOM produces the configuration program, that is used by the PRISMA-CASE tool (Cabedo et al., 2005) in order to configure the PRISMA software architecture by instancing the PRISMA types. These instances will configure the software architectures of our Product Line. Therefore our DSPL are the diagnosis systems of each one of the specific domains.

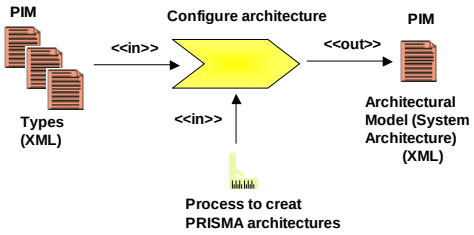


Figure 10: Configure the architectural model.

BOM uses the PRISMA-MODEL-COMPILER tool (Cabedo et al., 2005) to automatically generate the code (in .NET, C#) of the software architecture of the preceding task. The final diagnostic system, i.e. an instance of the DSPL, is executed on top of the PRISMANET middleware (Costa et al., 2005).

5 THE CASE STUDY

We have selected the field of infantile infectious diseases as an application domain from the field of medical diagnosis in order to know the requirements necessary to obtain the final software product in our Product Line.

The software system of infantile infectious diseases contains knowledge that uses a set of rules made of the signs and symptoms of diseases of the patient. These signs, symptoms and diseases have been provided by a paediatrician. The system proposed in this work, makes the medical diagnosis by introducing patient data into the system. This data is made up of sign and symptom values and is input by the final users of the system. As a result, the system obtains a diagnosis of the patient's disease. The final diagnosis is made from two types of diagnoses: clinical and laboratory. The objective of this is to provide a highly accurate diagnosis.

Medical diagnosis is understood as the process of the identification or recognition of a disease on the basis of the signs and symptoms (including laboratory studies) of the patient. The medical diagnosis represents the research process performed on the patient, and the diagnosis is based on the observations and reasoning of the doctor.

In medical diagnosis, the entity to be diagnosed is the patient, and the result is the disease that a patient has. The properties are signs and symptoms. These are classified in two abstraction levels: coarse grain and fine grain.

The process of medical diagnosis is the following: A syndrome is inferred from sign and symptom values of coarse grain. Two or more possible diseases are inferred from the syndrome. Deductive reasoning is applied in this part of the process. These hypotheses must be validated. A disease (validated hypothesis) is inferred from sign and symptom values of fine grain. Inductive reasoning is applied in this part of the process.

We present examples of the properties, rules, pre-hypotheses and hypotheses of the case study.

```
properties of level 0: cough, fever
properties of level 1: dry_cough,
constant_fever
pre-hypotheses: warth, parotiditis
hypotheses: pneumonia, bronchitis
rules:
IF (cough=true and fever=true and
  respiratory_dificulty=tue)
  THEN syndrome=warth
```

The variability sources or points in the diagnostic domain of our Product Line for this case study are the following.

- The hypotheses are inferred by means of different properties of the entities.
- The system exhibits a type of behaviour or reasoning strategy: differential reasoning. This reasoning strategy is the most widely used in solving medical diagnostic problems because it is suited to this kind of task.
- The entity properties of level 0 are the signs and symptoms of coarse grain (e.g. cough and fever). The entity properties of level 1 are the signs and symptoms of fine grain (e.g. dry cough and constant fever).
- The pre-hypotheses are syndromes (e.g. parotiditis) that are inferred by means of rules of level 1. The hypotheses are the diseases (e.g. pneumonia) that are inferred by means of rules of level 2. The diagnostic result is inferred by means of rules of level 3.
- In this case study, several hypotheses are generated. These must be validated in order to obtain only one validated hypothesis, which is the diagnostic result.
- The system offers the user the following functionalities (use cases): clinical diagnosis,

laboratory diagnosis, and the visualization of the results of the final diagnosis.

- The use cases are used by two final users (actors from the use case diagram): doctors or members of the laboratory. Two use cases can be invoked by the doctor, and only one use case can be invoked by a member of the laboratory. In this case study, there are three use cases and two actors, where one actor (the doctor) is associated to two use cases, and the other actor (the member of the laboratory) is associated to only one use case. Figure 11 shows the case use diagram for the medical diagnosis.

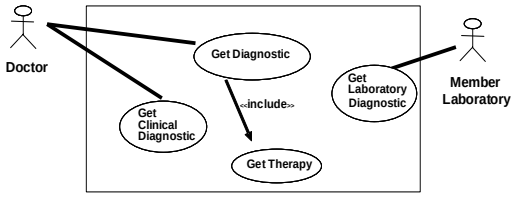


Figure 11: Use case diagram for a medical diagnosis.

The skeletons of the architectural elements are: Inference Motor (with three ports), Knowledge Base (with three ports), User Interface 1, i.e., the doctor (with two ports), User Interface 2, i.e., the member of laboratory (with one port), Diagnostic Connector 1 (with three ports), Diagnostic Connector 2 (with three ports), and Diagnostic Connector 3 (with three ports). The architectural model represents a product of our DSLP and is shown in Figure 12

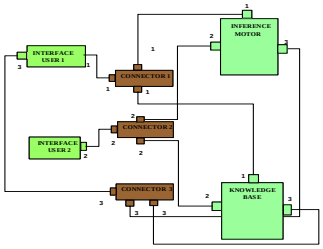


Figure 12: Visual metaphor of the architectural model of the medical diagnosis system.

We present the (partial) code that is automatically generated by the PRISMA-CASE tool. This code corresponds to the Knowledge Base Component of the case study.

```
namespace KBMD
{  [Serializable]
   public class
   KnowledgeBaseMedicalDiagnosis:
   ComponentBase
  { public class
      KnowledgeBaseMedicalDiagnosis
       string name:  base (name)
      { AddAspect (new FBaseMD ());
      InPorts.Add
      (“KnowledgeClinicalPort”,
      “IDomainMD”, “KNOWLEDGE_CLIN”);
      OutPorts.Add
      (“KnowledgeClinicalPort”,
      “IDomainMD”, “KNOWLEDGE_CLIN”);
      InPorts.Add
     (“KnowledgeLaboratoryPort”,
     “IDomainMD”, “KNOWLEDGE_LAB”);
      OutPorts.Add
     (“KnowledgeLaboratoryPort”,
     “IDomainMD”, “KNOWLEDGE_LAB”);
      InPorts.Add
     (“KnowledgeResultsPort”,
     “IDomainMD”, “KNOWLEDGE_RES”);
      OutPorts.Add
     (“KnowledgeResultsPort”,
     “IDomainMD”, “KNOWLEDGE_RES”);
}  }}
```

6 RELATED WORKS

There are a great number of works that are related to our approach. The methodologies and applications on this subject have produced a wide variety of research products, offering suggestions and solutions in specific domains.

A study made by (Liao, 2005) examines the methodologies of expert systems and classifies them into eleven categories. Two of these categories correspond to the systems based on rules and the systems based on knowledge. These categories have been taken into account in our work when using knowledge represented in the form of rules (Horn clauses) and facts (observable variables). Likewise, (Liao, 2005) mentions that the applications of the expert systems are built as specific domain problem oriented systems. In our work we present a case study in the medical diagnostic domain.

(Liao, 2005) also mentions that the development of expert systems has been characterized by the separation of knowledge and processes as independent units. In our architectural model, the elements in the type level are defined taking into account this concept, specifically when there is a component that contains the domain knowledge and another component that executes the inference process of the diagnosis.

(Giarratano et al., 2004) and other authors in the field of expert systems considered that the architectures of these systems are based only on components. The architecture of our system has integrated two approaches combining both

components and aspects. This increases the reusability and the maintenance of the system.

Expert systems have also been implemented in the development of different programming paradigms such as structured, logic, and object-oriented paradigms. These paradigms are oriented toward fourth-generation languages and visual programming methods to provide user-friendly communication. PRISMA provides an Architecture Description Language to define an architectural model that follows the MDA approach (http://omg.org.mda), which allows the automatic generation of code.

The integration of the DSBC and DSOA approaches are introduced by (Constantinides et al, 2000). The concerns and requirements described in this work are contemplated in our architectural model. The advantages of each one of these approaches are used to define the architectural elements with their aspects.

MDA proposes the definition of models at high abstraction level, which are independent of the technology (PIM). In our work, we have considered this line of research focusing on experts systems that are based on product lines.

Our work also applies the detection of the components based on the functional decomposition of the problem, which is compatible with the Architecture Based Design Method (ABD) methodology (Bachman, 2000). In ABD software architectures of the application domain are designed. This methodology has been applied in the building of our architectural model for medical diagnosis.

The work by (Garlan, 2001) is a very important reference in establishing the elements of a complex software system. In his work, he introduces the component, connector, system, input port, and output port concepts. These concepts have been included in our model.

Another work that is related to our approach is based on the contract concept of (Andrade et al., 1999). We have defined the connectors of the architectural models of the DSPL, incorporating the choreography concept in the connectors, which are specified by the protocol of the coordination aspect of the connectors.

There are many Architectural Definition Languages (ADL), which have advantages and disadvantages. This study has been done by (Medvidovic et al., 2000). The language proposed by (Loques et al., 2000) in their model R-RIO is the one that is the closed to the PRISMA-ADL. Their model has re-configuration capabilities like PRISMA; however their work does not incorporate the notion of aspect.

Software Product Lines have been an important discussion topic in the last decade. There are many works on this subject. Our research is related to the following works:

- (Batory et al., 2006) express the domain features in the Features Model, and they use Feature Oriented Programming as a technique for inserting the features.
- (González et al, 2006) applied the MDA proposal and Requirements Engineering for Product Lines.
- (Clements et al, 2002) use the SPL development approach, considering a division between domain engineering and application engineering for the reuse and the automation of the software processes.
- (Trujillo, 2007) has developed the XAK tool to insert features into XML documents by means of XSLT templates.
- (Ávila-García et al, 2006) has developed a MDA tool with functionalities of metamodeling over MOF and the transformations in ATC. The authors integrate the functionalities of the process modeling in SPEM, and RAS to package reusable assets .
- (Santos, 2007) proposes the development of a technique based on MDA for variability management in Software Product Lines.
- In (ACM, 2006) several works, related with to the Software Product Line Engineering have been published.

7 CONCLUSIONS

This paper presents BOM (Base-Line Oriented Modeling), wich is a framework that automatically generates diagnostic systems based on software product lines.

BOM has been designed to improve the development of diagnostic systems in following ways:

- To use the advantages of Expert Systems: incorporate several reasoning strategies in order to solve a problem by applying the most efficient one, and separate the inference process of the knowledge information from the application domain.
- To apply techniques from the field Software Product Lines by building a design that shares all the members of a program family. In this way, a specific design can be used in different products. Since we obtain a specific product from a series of previous models, the costs, time, effort, and complexity can be reduced.

- To construct Product Line Architectures in the PRISMA framework, is order to have the advantages of distributed systems, which will facilitate the management of complexity.
- To create an integrated and flexible approach to describe (medical) diagnosis architectural models that are complex, distributed, and re-usable by improving the development of expert systems for (medical) diagnosis following the PRISMA model (Pérez, 2006) to integrate the components and aspects.
- To apply MDA techniques to implement the systems on different platforms, and to automatically transform them and incorporate the features of the Features Model instances to obtain an executable application.

In the future, we want to extended the analysis of the diagnostic field in other application domains in order to increase variability and our Base-Line. Our Products Line will be able to offer more products. In addition, we plan to validate our approach in other case studies, and compare the performances of the generated Expert Systems with other obtained using other approaches.

ACKNOWLEDGEMENTS

This work has been funded under the Models, Environments, Transformations, and Applications: META project TIN20006-15175-605-01.

REFERENCES

Andrade L. and Fiadeiro J., 1999. Interconnecting Objects via Contracts. OOPSLA´99.

Ávila-García O., García A. E., Rebull V. S., y García J. L. R., 2006. Integrando modelos de procesos y activos reutilizables en una herramienta MDA, *en XI Jornadas de Ingeniería de Software y Bases de Datos JISBD'2006*, Barcelona, España.

Bachman F., Bass L., Chastek G., Donohoe P. and Peruzzi F., 2000. The Architecture Based Design Method. *Technical Report CMU/SEI-2000-TR-001, Carnegie Mellon University*, USA.

Batory D., Benavides D., and Ruiz-Cortés A., 2006. Automated Analyses of Feature Models: Challenges Ahead. *ACM on Software Product Lines*.

Cabedo R., Pérez J., Carsí J.A. y Ramos I., 2005. "Modelado y Generación de Arquitecturas PRISMA con DSL Tools", *en Actas del IV Workshop DYNAMICA*, Archena, Murcia, España.

Clements P. and Northrop L.M., 2002. *Software Product Lines: Practices and Patterns.* SEI Series in Software Engineering, Addison Wesley.

Constantinides C.A., and Errad T., 2000. On the Requeriments for Concurrent Software Architectures to Support Advanced Separation of Concerns. In Proceedings of The OOPSLA 2000, *Workshop on Advanced Separation of Concerns in Object-Oriented Systems.*

Costa C., Pérez J., Ali N., Carsí J.A. y Ramos I., 2005. "PRISMANET: Middleware: Soporte a la Evolución Dinámica de Arquitecturas Software Orientadas a Aspectos", *en Actas de las X Jornadas de Ingeniería del Software y Bases de Datos*, Granada, España.

Czarnecki K., and Eisenecker U., 2000. *Generative Programming: Methods, Tools, and Applications.* Addisson-Wesley. ISBN 0-201-30977-7.

Garlan D., Cheng S. and Kompanek A. J., 2001. *Reconciling the Needs of Architectural Description with Object Modeling Notations*. Science of Computer Programming Journal, Special UML Edition, Elsevier Science.

Giarratano, J., and Riley, G., 2004. *Expert Systems: Principles and Programming.* Fourth Edition: (Hardcover), ISBN: 0534384471.

González-Baixauli B. y Laguna M. A., 2005. MDA e Ingeniería de Requisitos para Líneas de Producto. *Taller sobre Desarrollo Dirigido por Modelos. MDA y Aplicaciones. (DSDM´05*), Granada, España.

Greenfield J., Short K., Cook S, Kent S., and Crupi J., 20004. Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools. Wiley.

Liao S.-H., 20005. "Expert Systems Methodologies and Applications- a Decade Review from 1995-2004", *in Expert Systems with Applications*, Vol. 28, Issue 1.

Loques O., Sztajnberg A., Leite J., and Lobosco M., 2000. On the Integration of Meta-level Programming and Configuration Programming. *In Reflextion and Software Engineering (special edition*), Lectures Notes in Computer Science, Springer-Verlag, Heidelberg, Germany

Medvidovic N., and Taylor R.N., 2002. A Classification and Comparison Framework for Software Architecture, *in Proceedings of IDEAS*, Cuba.

Pérez J., 2006. *PRISMA: Aspect-Oriented Software Architectures*. PhD. Thesis of Philosophy in Computer Science, Polytechnic University of Valencia, Spain.

Santos A.L., Koskimies K., and Lopes A., 2005. *Using Model-Driven Architecture for Variability Management in Software Product Lines*. Ph Thesis Proposal Facultade de Ciencias de la Universidade de Lisboa, Portugal.

Software Product Line Engineering Communications of the ACM. 2006, Vol 49, Number 12, pp 28-88.

Szyperski C., 1998. "*Component software: beyond object-oriented programming"*, ACM Press and Addison Wesley, New York, USA.

Trujillo S., 2007. *Feature Oriented Model Driven Product Lines.* PhD. Thesis, The University of the Basque Country, San Sebastian, Spain.

INTEGRATION SOLUTION FOR THE ACCESS TO HETEROGENEOUS MEDICAL DEVICES
Communication with Healthcare Devices in Intensive Care Units

Susana Martin-Toral, Jose Luis Rodriguez-Gonzalez
Computer and Information Technologies Division. CARTIF Foundation
(Centre for Automation, Robotics and Information and Manufacturing Technology)
Ave. Francisco Valles 204-205, Parque Tecnologico de Boecillo, Valladolid, Spain
susmar@cartif.es, josrod@cartif.es

Javier Perez-Turiel
Head of Biomedical Engineering Division, CARTIF Foundation
javper@cartif.es

Keywords: Medical devices, drivers, VITAL (Vital Signs Information Representation), manager/agent architecture, MDIB (Medical Data Information Base).

Abstract: This paper presents a free Critical Care Information System (CCIS) that shows an essential infrastructure for critical care medical and nursing practice. Specifically, a Patient Integral Analysis Aid System (SAIP) in Intensive Care Units (ICU) has been developed to cover the needs discovered in these scenarios. An importart part of this system is related to medical equipment, that offers important information to help in medical diagnosis. ICU patients are usually connected to several of these devices which register their physiological parameters. The integration of these devices, in order to exchange the generated information, is difficult because they are developed by different manufacturers and with different communication protocols and information representations. Due to this, it has been necessary to develop a set of communication drivers for each medical device, according to the current regulations. To reach this objective, the developed drivers have a common interface for the access and collection of medical device data. The main goal of the present paper is to show the work done to obtain a real interoperability among medical devices from different manufacturers and with different communication protocols in ICU services for automatic data collection, storage and retrieval.

1 INTRODUCTION

In this paper a free Critical Care Information System (CCIS) is presented. CCIS is an essential infrastructure for critical care medical and nursing practice. This premise is described in terms of the critical care environment, clinical decision making in critical care, increasing demands for information about quality and costs, and national initiatives for the sharing of healthcare information. This kind of systems are designed to collect, store, organize, retrieve, and manipulate all the data related to the care of the patient in Intensive Care Units (ICU). Some of the main purposes of these systems are: automated vital signs capture, cardiac rhythm analysis and dysrhythmia detection, reporting laboratory results, entry and transmission of physician orders, admission, discharge and transfer of patients, calculation of medication doses, fluid infusion rates, shift, and daily intake and output volumes, organization of patient records, calculation of patient plan of care, entry and organization of care documentation and prompting the caregiver for recorded documentation (Adhikari and Lapinsky, 2003; Fraenkel et al., 2003; Frize et al., 1997). In this research project a Patient Integral Analysis Aid System (SAIP) in Intensive Care Units (ICU) has been developed, which is now being used in the ICU of the University Clinical Hospital of Valladolid, in Spain.

The SAIP is a system based on free software GNU GPL (General Public License) (Ope, 2007) composed by a set of computer applications to facilitate and improve the attention received by the patient in intensive care units (Martin et al., 2004; Jose, 2007). It provides a support system to collect, store and manage the clinical information in the ICU, with the purpose of facilitating the daily medical routine and to improve the quality of the attention provided to the patient by overcoming the actual difficulties of inter-

connection and storage (Shortliffe, 2001).

The ICU patient is, habitually, connected to one or several devices - called in this paper medical devices - that allow information on several of their physiological parameters to be measured, as shown in figure 1 with a vital signs monitor. These devices generate a great amount of information, but their capacity of storage is limited. In addition, they are usually from different manufacturers and models, and they are not interconnected with the rest of the devices. That is why the integration of the generated information is difficult. This heterogeneity of medical devices is one of the main problems for making an integral system for the ICU. Each device has its own communication protocol and its own connection interface, for example, RS-232, Ethernet, MIB (Medical Information Bus), etc. and there is no interface or protocol that allow the unification of communication and access between different devices and a user application that wants to communicate with them. In order to solve this problem, the developed system unifies, for each type of medical device, the data obtained from the different medical devices that compose the system, by using diverse standards of denomination and storage.

Figure 1: Monitor connected to the ICU room PC.

The initial situation in the ICU consists of a series of medical devices, such as infusion pumps, respiratory ventilators and vital signs monitors, each with a different type of connection, a different communication protocol and without a common interface that allows homogeneous access to the data that these devices provide (Martin et al., 2004; Jose, 2007). Due to this, it has been necessary to develop a set of communication drivers for each medical device that allow the information provided by the medical devices be collected and stored automatically, although these are from different manufacturers, as well as the information contributed by the medical staff. Because of that, the drivers developed for the SAIP system have a common interface for the access and collection of data from the medical devices.

Although the main objective of the present paper is to show the work done to obtain the homogeneous interoperability between medical devices in the ICU, additional objectives of the SAIP system (Jose, 2007) are summarized as follows:

- Automation of the medical workflow: electronic management of documents.
- Collection and storage of other types of manually acquired information: medical devices without possibility of automatic capture of information, connectionless devices and data from the patient exploration task filled in directly by medical staff.
- Calculation of medical statistical indexes: APACHE, NEMS,etc.
- Consultation of registered information. The information stored in the system is readily accessible.
- Document elaboration of patients' management (registries, personal information, etc).
- The SAIP must comply with the National medical information security laws.
- Interchange of data with other hospital services. The intensive care unit is not an isolated service, so it needs constant communication with other hospital services. The efficient implementation of defined interfaces should ensure communications between the diverse systems/services of the hospital.
- To manage a safe and trustworthy access to the system. Most of the information generated in hospitals is personal medical information and therefore is confidential.
- Development under the concept of free software (GPL license). The development of applications based on free software is an area of activity of ever growing interest that generates numerous expectations, given the ample possibilities it offers.

The rest of this paper is organized as follows : section 2 describes the proposed control architecture of the SAIP system in the ICU, paying special attention to the part related to the integration of medical devices. Section 3 describes the communication drivers developed to allow the medical devices to communicate with the SAIP System. In addition, it describes the dynamic libraries of plugin type solution proposed. Section 4 describes the agent entity whose function is to communicate with the different medical devices with its associated database. This database stores the device data. In section 5 the results of the

system development are presented. Finally, a summary of the work done and some conclusions are included in section 6.

2 ARCHITECTURE SOLUTION

The SAIP is a system based on the use of computer, control and communication technologies to generate a set of computer applications whose main objective is to ease and improve the attention received by patients in intensive medical services. From a physical point of view, the planned system requires the connection of medical devices (vital signs monitors, ventilators, infusion pumps, etc.) to external elements, the installation of computer systems and communication networks, and the development of software applications for data recording, storage, processing and access.

The proposed system architecture is based on a manager/agent framework, and in general it can be seen as a three layer architecture (data, business logic and presentation). Furthermore, the system functionality has been divided in two different parts: the integration environment and the execution environment or healthcare information system (HIS) (Scherrer and Spahni, 1999; Martin et al., 2004).

In general, the aim of the *integration environment* is to unify the information obtained from every kind of system device (both medical and computer). To achieve this objective, several network device manager frameworks have been used (Leinwand and Fang, 1995), such as the ones based on the OSI (Open System Interconnection) interconnection basic reference model or the ones based on SNMP (Simple Network Management Protocol) (Stallings, 1999; Case et al., 1990). On the other hand, the *execution environment*, or HIS environment, is the one in charge of implementing the business logic, so it develops the thickness functionality provided by the SAIP. The modules that form the HIS environment are located throughout the diverse physical devices that house the application, based on the information type they handle. This information is provided by both the integration environment and users, or through other applications and external systems (admission information, laboratory results, etc.). Therefore, the execution environment has the elements needed to process the information and to provide the functionality required by users.

From this architecture, the present work emphasizes the part dedicated to the *communication and information processing provided by medical devices*, according to norms (CEN/CENELEC, 1993; CEN, 2000). The goal pursued in this aspect is to have a system that allows the interaction with patient vital support devices, so as to make the data and alarms generated by these devices available for users (doctors, nursing staff, nursing assistant, etc.). Since these medical devices are in the patients' emergency rooms, the computer equipment entrusted to interact with them must also be located in the same emergency rooms. This IT equipment (pesonal computer) is called, in our system, *ICU room PC*. The ICU room PC architecture is shown in figure 2, and its main characteristics will be detailed in the rest of the section.

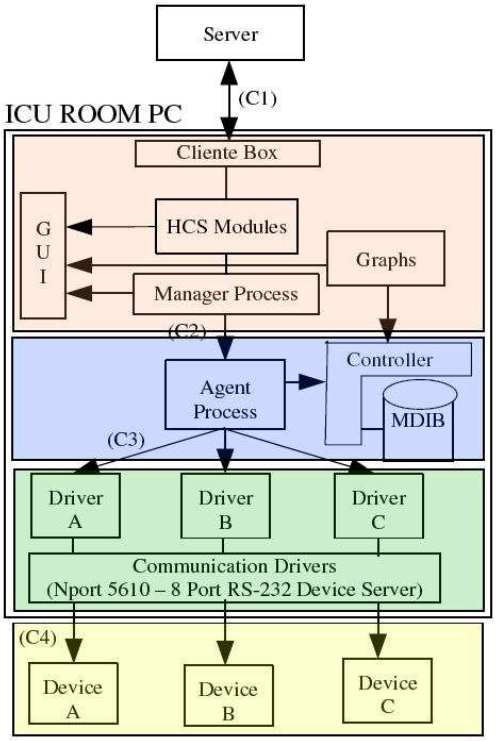


Figure 2: Block diagram of ICU Room PC.

For the ICU room PC connection architecture, the main objective of the integration environment is to unify, for every type of medical device and as far as possible, the data obtained from the different devices that compose the system. This environment, therefore, converts and translates the hardware dependent information provided by medical devices into an annotation and nomenclature understandable by all the devices. Moreover, it also unifies the way in which the calls are made to access information. In this way an integrated and homogeneous medical device management is obtained.

With these objectives, the integration environment uses some parts of the VITAL norm (Vital Signs Information Representation) (CEN, 2002) to define the

semantics of the information provided by medical devices as a structured set of objects and methods to represent medical information. The VITAL experimental norm provides the definition of an independent common device representation of vital signs information, and the definition of a common model for accessing to this information. It means that VITAL looks for real interoperability between medical devices even if they are from different manufacturers (CEN, 2002).

In general, the information provided by medical devices is transformed into VITAL nomenclature to be stored in a database called MDIB (Medical Data Information Base). An agent entity is responsible for updating this database, which in addition attends to the requests made by a manager entity, as can be seen in detail in section 4. In short, the integration environment needs:

- A set of drivers that allow communication with diverse medical devices. Each one is device dependent, so they must speak the same language as the device they access.
- The translation of device dependent information into VITAL nomenclature.
- An agent entity that gathers this information periodically.
- A database where information is stored.

In figure 2 the connection schema of the elements housed in the ICU room PC is shown. Three differentiated groups are distinguished in this schema. From the lower level of abstraction to the upper, firstly the set of drivers can be seen, then the agent entity and its related database, and finally the business logic and the graphic user interface housed in the ICU room PC. Each of these groups is presented in detail in the following sections.

3 COMMUNICATION DRIVERS

Drivers are programs in charge of serving as middlewares between the computer operating system and the different elements connected to it. In this particular case, the system works with two different kinds of drivers, as can be seen in figure 2. The first one fits the definition given before, and is called “communication driver” in the figure. This driver allows the communication with medical devices through a communication element. Currently, the system works with a unique communication driver, that is, a serial port hub of Moxa trademark (Moxa’s NPort 5610 8 Port RS-232 Device Server). The reason is that almost all the medical devices provide their output through an RS-232 interface, so we have chosen to work with this type of device server, that allows a greater the number of RS-232 interfaces than a conventional computer. Perhaps in the future the system could need other types of drivers for medical device communication, maybe through Ethernet or MIB (Medical Information Bus) interfaces.

The second driver type presented in the schema is related to special programs developed to implement the communication protocol of each medical device. Therefore, there must be a driver for each type and/or model of medical device. These drivers can be seen as translator gateways, since they translate data requests from the agent entity, who makes them in a homogeneous way, into a language understandable by the corresponding medical device, and then they respond in a standard VITAL format. In this way the agent entity can talk to the different devices using the same language, and without worrying about which device it is talking to. The main objectives of these drivers are:

- To implement the communication through the serial port (extendable to other communication interfaces).
- To implement the proprietary communication protocol of each medical device.
- To attend to manager entity requests.
- To translate the information provided by devices into VITAL format.

In this work, a plugin is considered a driver implemented for every type or model of medical device (the second type of driver seen before), since its compilation is made as plugin libraries that can be loaded later by an agent entity. These plugins have been developed using the same interface, so the agent can call their functions in a homogeneous way and without the need to know the device with which it works, or from which manufacturer it is. In this way, the system is totally scalable in order to increase the number of medical devices to work with, and it ensures the communication with heterogeneous devices from diverse models and manufacturers.

When the incorporation of a new medical device is needed in the ICU, and whenever the new device presents some external communication port and protocol, it will be enough to implement the appropriate driver (filling the specified interface and VITAL representation), to generate the plugin library and to inform to the agent entity about the new plugin name and location. Once the plugin is incorporated into the system, the agent entity will be able to communicate and extract information from this device.

4 AGENT ENTITY AND ITS RELATED DATABASE

The agent entity is the element in charge of the communication with the different medical devices (connected to the ICU room PC by the serial ports hub) through the corresponding plugins. Its main target is to collect the information as well as the alarms provided by the different medical devices, to be monitored and validated. Every agent entity communicates with the medical devices assigned to a specific patient and, with the data provided by them, must fill in the database named MDIB, and must respond the requests from a manager entity. Therefore at this level it is a manager/agent architecture.

The main functionalities of the agent entity are:

- To detect the connection of a new medical device in the serial port hub assigned to one ICU room and one patient. This detection is carried out by the own hub, which is able to send an SNMP trap to the ICU room PC indicating the status changes in DCD and DSR signals in each hub's port.
- To identify the connected medical device, that is, to detect the type of device the agent entity must work with to use the suitable plugin. With this aim, the agent entity loads every plugin and checks if it exists an appropriated communication with the device related to the plugin. If the device responds its plugin will have been detected, and therefore the agent entity will know the type of device which it works with.
- The previous items provide the system with a plug-and-play functionality related to the connection of medical devices to the system. This functionality simplyfies the connection of these medical devices to the nursing staff.
- To collect periodically the data provided by these devices, and update the MDIB with them.
- To detect alarm conditions in the devices using pull or push methods, according to the behavior of each one.
- To notify the detected alarms to the adecuated element of the systems for its processing and broadcast.
- To attend the requests of the manager entity, that must ask for and provide the information from medical devices to the business logic layer, to be processed and showed to system users.

The MDIB, database containing the information provided by medical devices, has an information model according with a part of the structure and nomenclature of the VITAL norm. This part is used to represent the different information objects. Either the agent entity, or other modules and elements, could need to access to the MDIB database.

5 MAIN RESULTS

In order to understand the obtained results it has to be considered the scenario where the project has been developed:

- Each ICU room has a ICU room PC.
- An ICU room can have more than one patient.
- For each patient there is an agent entity and a MDIB associated to it. Therefore in each ICU room PC there will be so many agent entities and MDIB as patients are in the ICU room.
- All the data gathered during a patient's stay in the ICU is stored in the MDIB.
- This MDIB is filled in a periodical way. In every period the data collected by the drivers is used to fill the database.
- When a patient leaves the service, and therefore the system, the data stored in the MDIB relatec to that patient is erased.

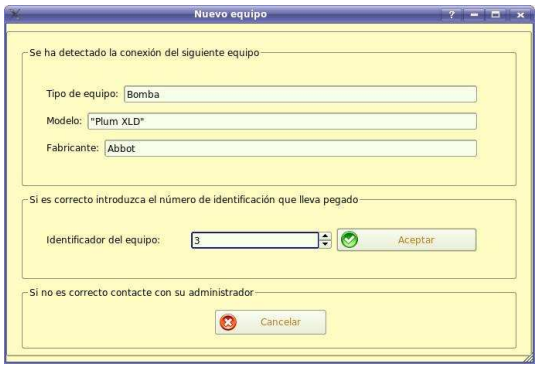


Figure 3: Device identification screen.

- The agent entity must be always active, waiting for the connection/disconnection of a medical device, even if there is no patient in the room.
- When there is no patient in the room and a device is connected, the agent assumes that there is a new patient in the room.
- Whenever a device is connected, if the agent is able to recognize it, a screen appears indicating the device detected (type of device, manufacturer, etc.). If the displayed data is correct, the user must introduce the number that identifies that device univocally, as it is shown in figure 3.

- The agent monitors each device using its sampling period. The data is stored periodically in the MDIB of the ICU room PC, whereas the detected alarms are sent to the appropriate process to be handled. An example of an alarm originated by an infusion pump is shown in the next listing (taken from the standard output):

```
Hilo alarmas: Se ha detectado 1 nueva
              alarma en el equipo 3
---------------------------------------
Identificador de dispositivo: 3
---> Hilo alarmas: ALARMA 0:
        Equipo: Bomba
        Modelo: "Plum XLD"
        Descripcion: Fallo en cassette.
        Tipo de alarma: tecnica
        Prioridad: baja
        Fecha de produccion: 22/06/2007
        Hora de produccion: 08:14:32
---------------------------------------
```

- Once the data is stored in the MDIB, the medical staff has the possibility of visualizing it as graphs (trend curves) corresponding to different physiological signs from the medical devices. Two examples of these trend curves can be seen in figures 4. and 5. Only the vital signs with a green spot are showed in the graph.

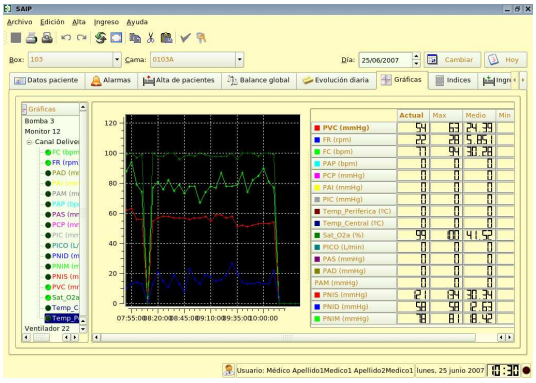


Figure 4: Trend curves from patient monitor data.

- When the patient leaves the service all the medical devices must be disconnected, while the agent entity waits for the arrival of a new patient.

During the project, the following medical devices communication drivers have been developed, tested and included into the system:

- Lifecare XL Infusion Pump from Abbott Laboratories.
- Evita Ventilator from Drager Medical.
- Vital signs monitor from Agilent Technologies.

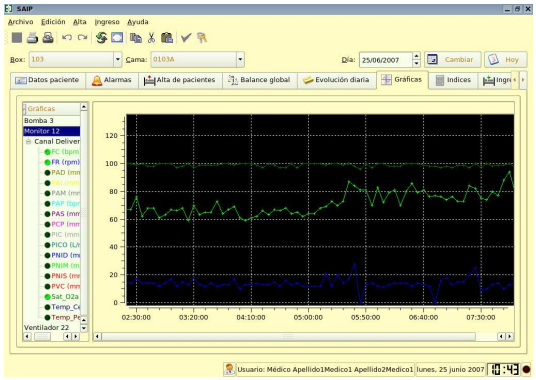


Figure 5: Detail of trend curves.

The software developed in this project has been implemented using the C++ language and the Trolltech's QT library (Tro, 2007).

6 CONCLUSIONS

In this paper it has been presented the components of the SAIP system developed to interact with medical devices (vital signs monitors, ventilators, infusion pumps, etc.) in charge of the patient's vital support in Intensive Cares Units (ICU). The details of the components of the ICU room PC architecture, being part of the global SAIP system, have been shown. It has been necessary to develop communications drivers with a common interface to facilitate the connection to the different devices in a uniform way, and the acquisition of the data provided by the medical devices following the indications of the medical staff responsible for the ICU. These communication drivers have been developed as plugin libraries to allow the SAIP application to manage them through an agent entity, while only those corresponding to medical devices connected to the patient are loaded and used.

On the other hand, once the data of the medical constants of each device are gathered and transformed into VITAL nomenclature, they are stored in the MDIB of the corresponding ICU room PC. This allows to visualize as graphs (trend curves) the values stored for the different medical devices. These graphs allow the medical staff to see the evolution of the different medical signs provided by the devices during a complete cycle of 24 hours, or the ones stored for any day of the stay of the patient in the ICU.

Considering the implementation and the behaviour of this part of the system, the benefits and remarkable advantages related to the handling of information provided by the medical devices are:

- Automatic and periodical collection of values of

vital signs, avoiding the manual introduction of these values in the system.

- Automatic collection of the alarms produced by these devices and diffusion for the knowledge of the medical staff of the ICU service.
- Storage of all this information. Capacity to consult data of different stages from the beginning of the stay of a patient. Usually medical devices do not have the capacity to store the information generated during all the stay of the patient in the ICU. Some of them have buffers of storage, but they are not big enough to store all the information relative to the complete entrance of the patient.
- Plug-and-play functionality, that simplifies the use of the system for communicating with the medical devices, specially for the nursing staff.
- Automatic collection of the constants indicated in the treatment, always with the corresponding validation of the nursing staff. Obtaining of nursing reports relative to these constants.
- Taking of values for semi-automatic generation of balances.
- Generation of trend curves.
- Compilation of the information in a unique and standard format that facilitates its integration and handling and provides a real interoperability among devices.
- An application that facilitates a joint visualization for the generation of diagnoses with all the information of several medical devices.

As it can be seen, the main purpose of the system is to facilitate the collection, storage and subsequent processing of the information provided by the medical devices connected to patients in the intensive care units in an homogeneous and standard way. The automation of the collected data has been possible to make it available to the medical staff in a friendly way, and a real interoperability among heterogeneous medial devices has been obtained.

ACKNOWLEDGEMENTS

This work has been funded by the SACYL (Sanidad Asistencial de Castilla y Leon - Health Welfare of Castilla y Leon), through agreement of collaboration between the University Clinical Hospital (HCU) of Valladolid and the Centre for Automation, Robotics and Information and Manufacturing Technology (CARTIF Foundation). We appreciate the collaboration of the intensive medicine staff of the HCU and the CARTIF development team involved in the project.

REFERENCES

(2007). Open soruce home page. URL: http://www.opensource.org. Last visit: june 2007.

(2007). Trolltech home page. URL: http://trolltech.com/. Last visit: june 2007.

Adhikari, N. and Lapinsky, S. E. (2003). Medical informatics in the intensive care unit: overview of technology assessment. *Journal of Critical Care*, 18(1):41–47.

Case, J., Fedor, M., Schoffstall, M., and and, J. D. (1990). RFC 1157 - Simple Network Management Protocol (SNMP). Technical report, Network Working Group. URL: http://www.faqs.org/rfcs/rfc1157.html. Last visit: june 2007.

CEN (2000). ENV 13735:2000. Health informatics - Interoperability of patient connected medical devices. European prestandard. English version.

CEN (2002). CEN/ISO ENV13734. Health informatics. Vital signs information representation. European prestandard.

CEN/CENELEC (1993). IEC 60601-1. Medical electrical equipment. General requierements for safety. European standard.

Fraenkel, D. J., Cowie, M., and Daley, P. (2003). Quality benefits of an intensive care clinical information system. *Critical Care Medicine*, 31(1):120–125.

Frize, M., Trigg, H., Stevenson, M., and Solven, F. G. (1997). Decision-support systems for critical care. In *American Medical Informatics Association Conference*.

Jose, S. S. (2007). Modelado y desarrollo en software libre de un sistema de analisis integral de pacientes en unidades de cuidados intensivos. In *FSWC 2007: Free Software World Conference 3.0*.

Leinwand, A. and Fang, K. (1995). *Network Management, a practical perspective*. Addison Wesley, London, 2nd edition.

Martin, S., Gutierrez, S., Jose, S. S., and Barrientos, F. (2004). Sistema de ayuda al analisis integral de paciente en unidades de cuidados intensivos (SAIP). In *INFORMED 2004, X Congreso Nacional de Informatica Medica. El tratamiento de la informacion, instrumento de valor aadido en la practica medica*, pages 185–192. SEIS, Sociedad Espaola de Informatica de la Salud.

Scherrer, J. R. and Spahni, S. (1999). Healthcare information system architecture (HISA) and its middleware models. In *AMIA Symposium, 1999 - amia.org*. AMIA: American Medical Informatics Association.

Shortliffe, E. H. (2001). *Medical Informatics. Computer Applications in Health Care and Biomedicine*. Springer, New York, 2nd edition.

Stallings, W. (1999). *SNMP, SNMPv2, SNMPv3, and RMON1 and 2*. Addison Wesley, London, 3th edition.

ELECTROPHYSIOLOGICAL CONTROL SIGNALS FOR PERSONS WITH NEURODEGENERATIVE CONDITIONS: BLENDED CONTROL SIGNALS

Ana Londral
ANDITEC, Lisbon, Portugal
analondral@anditec.pt

Luis Azevedo
Centro de Análise e Processamento de Sinal, Instituto Superior Técnico, Lisbon, Portugal
pclma@ist.utl.pt

Pedro Encarnação
Faculdade de Engenharia, Universidade Católica Portuguesa, Lisbon, Portugal
pme@fe.ucp.pt

Keywords: Electrophysiological control signals, Assistive Technologies, User Interface, selection methods, progressive conditions.

Abstract: Severe neurological conditions may considerably affect one's functional capabilities. Special computer interfaces and access methods have been developed in attempt to provide a mean to overcome the functional disabilities experienced by persons in these conditions. In this paper, a case study on the usage of a brain-body interface by a young man with Amyotrophic Lateral Sclerosis is presented. From the study different ways of interacting with the computer, beyond the traditional direct selection and scanning methods, emerge. These resort to control signals that combine binary and continuous features, *blended control signals*. Such control signals may provide more flexible and efficient ways of interacting with Assistive Technology systems, especially for those individuals with neurodegenerative conditions.

1 INTRODUCTION

Various diseases or conditions may impose severe limitations in one's motor abilities and consequently lead to communication disorders. These diseases and conditions can be divided in *progressive* and *static or improving* (Glennen and DeCoste, 1996). Examples of progressive conditions are neurodegenerative diseases, as Amyotrophic Lateral Sclerosis (ALS), Multiple Sclerosis or Parkinson, and some oncological conditions. Brainstem strokes, traumatic brain injuries or spinal injuries are included in static or improving conditions, as they remain unchanged or improve over time.

Assistive Technologies can be defined by "*any item, piece of equipment or product system whether acquired commercially off the shelf, modified, or customized that is used to increase or improve functional capabilities of individuals with disabilities*." (United States Congress, 1998). Although there are many definitions for AT, *the main objective of assistive technologies (equipments and services) is to contribute to a better quality of life of the many persons affected by disabilities worldwide, through the integration of technological aspects in equipments, services and contexts* (Azevedo, 2006).

This paper is focused in AT systems, based on computer interaction, for persons with neurodegenerative conditions, i.e. progressive conditions caused by neurodegenerative diseases. In this context, individuals experience progressive decline in motor functioning, which dramatically affects their quality of life. Neurogenic communication disorders are a common consequence of neurodegenerative conditions as individuals progressively loose their ability to write and/or speak. Through computer interaction these persons may access to communication aids for writing or speaking.

AT selection for people with neudegenerative conditions is a big challenge since the progression of the disease must be "previewed", as well as other factors related to the individual's context. The progression of these diseases will lead to different needs and capabilities along the different stages of the disease. Flexibility is thus an utmost important characteristic for AT systems, which have to respond to individuals needs during all stages and conditions.

Considering AT systems based on computer interaction, user interface is an important part of the system, which translates users input signals into control signals. The most common user interfaces for severe neurodegenerative conditions are the ones using eyetracking techniques and the ones based on electrophysiological signals (Felzer and Nordmann, 2006). User interfaces are much dependent on the input signals that the user can control. The problem of the type of electrophysiological control signals that persons with neurodegenerative conditions can generate to access to AT devices, from early to late stages of disease, is addressed in this paper.

The paper is organized as follows. In Section 2, a brief description of user interfaces and typical selection methods used in AT systems context are exposed. The use of electrophysiological signals as control signals for AT systems is addressed in Section 3. Section 4 contains a description of a case study, which aims at evaluating the use of a brainbody interface by a young man with ALS to access to a computer as a communication device. This case study is discussed in Section 5 stressing the types of control signals the user was able to generate and proposing a new class of control signals – *blended control signals*. Paper conclusions are presented in Section 6.

2 USER INTERFACES

One of the critical elements of AT systems for persons with neurodegenerative conditions is the User Interface (UI). The UI receives user's input and translates it into control signals to access to the AT devices. These signals can be generated by various movements, such as hands, eyes or head movements, or even by other body sources as, for example, electroencephalograph signals. Control signals are then very dependent and conditioned by user's physical and context conditions.

A general representation of a UI for AT Systems is proposed in (Cook and Hussey, 2002) as shown in Figure 1.

Figure 1: User Interface of an AT System.

The Selection Method defines the way the user will select each element of selection set. Typically, AT devices provide Scanning or Direct Selection methods. Direct selection is possible if the user can generate at least as many control signals as the selection set. Otherwise, user has to resort to an indirect selection method (e.g. scanning) to pick an element of the selection set.

For example, given the task of writing in a computer, one may use a direct selection method pressing each key on the keyboard (AT device); however, if the person is not able to directly select each key, she needs to use a scanning method controlling it with one or more binary signals. Scanning method is much slower that direct selection method. However, there are many strategies that try to make this selection method more efficient according to users' abilities (Cook and Hussey, 2002).

This traditional strict division of selection methods ignores the possibility of having other kinds of interaction, based on control signals richer than simple on/off signals though not rich enough to control a 2-axis interaction (as showed in Figure 2 for the example of access to a virtual keyboard).

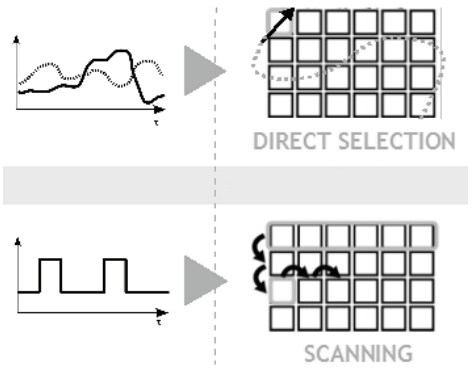


Figure 2: Example of access to a virtual keyboard. Traditional division for selection methods consider direct selection (continuous control signals for 2-axis control) or scanning method (based on one or two binary control signals).

When focusing on progressive conditions, AT systems must consider different kinds of access, being flexible to adapt to users' functionality. In this paper, the search for other kind of selection methods, based on electrophysiological control signals is discussed.

3 ELECTROPHYSIOLOGICAL CONTROL SIGNALS

Technology development in the field of biosensors has shown that individuals can generate and control various kinds of output signals that can therefore be used as control signals. In particular, todays control signals that are generated within the individual body can be used for man-machine interface.

When evaluating a person in a later stage of a neurodegenerative condition, often the main problem is to find *one* control signal that the user is able to intentionally generate. Even one single control signal supports an indirect access method, allowing a selection within a given set. The use of electrophysiological signals brings new perspectives on the number and type of control signals that a user with severe neurodegenerative conditions may generate.

At the skin surface level, two different types of signals can be captured: electric (e.g. electromyography, electrocardiography) and non-electric information (e.g. temperature, blood pressure) (Allanson, 2002). Typically, the former are the ones used for AT control systems, as it is the case of the AT system presented in this paper.

In case of individuals with neurodegenerative diseases, especially in later stages of the disease, these signals can be an efficient way of generating control signals. For example, an individual with very low motor control, who can't press a switch, can be able to generate control signals captured by an EMG sensor. In fact, an electrophysiological signal can provide a motor independent control signal even for persons in *locked in* state (Wolpaw et al., 2002; Wills and MacKay, 2006). However, an important issue to consider is that, due to its physiologic nature, electrophysiological information depends on the physical and environment conditions of each individual (such as diseases, fatigue, humour, environment temperature, familiar context, etc.). Thus, it is important to know the physiological mechanisms that produce the signals and how these signals are affected by referred conditions.

Therefore, in AT systems design, each case is a singular case, influenced by individuals' unique conditions and particular disease progression.

4 CASE-STUDY

A *Small Number Design* methodology (Iacono, 1992)(Stevens and Edwards, 1996) was used in order to evaluate the interaction of an individual with ALS with a brainbody interface (™Brainfingers Cyberlink). This brainbody interface consists on a headband with three surface electrodes placed on the forehead. The control signals generated by this interface are based on muscle and brain potentials, and are called *brainfingers* (Junker, 1995).

The individual that voluntarily participated in the study is in a later stage of the disease for some years. He can control very few movements and uses a pressing switch activated by slight head movements as the control interface to his communication aid. He is thus able to control a scanning process in software *The Grid*© for communication purposes and Internet access. With this system, this person wrote a published poetry book.

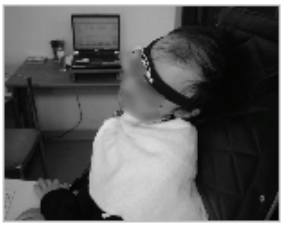


Figure 3: User studied using ™Brainfingers Cyberlink interface in a training session.

The motivation for this case study was twofold: are there alternative ways (and more efficient) for this individual to interact with an AT system?; is it possible for him (using ™Brainfingers Cyberlink interface) to generate more control signals or "richer" than binary control signals?

4.1 Test Design

A protocol for evaluation was developed and tested aiming at studying the control signals that the user was able to generate with the interface. Starting from the binary signal that the user used before, "richer" signals where progressively attempted. The tests followed the four steps described below.

a) One binary control signal

To gain confidence with the system, the user was firstly asked to use the AT system by means of his pressure switch, as he is used to. Then, the mechanical switch was replaced by the brainbody interface. Different sources of muscle potencial were essayed as a binary control. The signal generated by opening the jaw was found to be more efficient. In fact, this is the gesture that user does to communicate to his close friend and physiotherapist as a 'yes'.

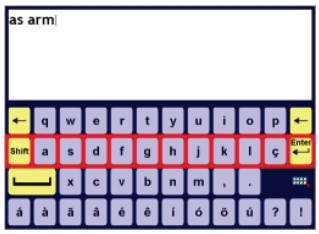


Figure 4: Virtual keyboard used to evaluate interaction using a binary control signal to write a sentence by a scanning method.

b) One continuous control signal

After getting used to access to computer using a binary *muscle signal*, the user was challenged to play a game where he had to move a bar in one axis to catch a ball. The bar could be controlled by user regulation of the *muscle signal* amplitude.

c) One "continuous and discrete" control signal

After being able to generate one *binary* control signal and one *continuous* control signal, the user was asked to access to his communication software using the combination of these two control signals. For that, a special one-row keyboard was designed (see Figure 5) and the user had to select each cell in a specific order. In order to do that, a continuous signal had to be controlled between two thresholds to move the selection bar. When this bar is in the desired position, the signal should be raised above the second threshold, thus making the selection (see Figure 4). In the designed application, the continuous signal amplitude within the two thresholds controlled the movement within the row, and the second threshold was used to generate a binary control signal for selection of the highlighted cell.

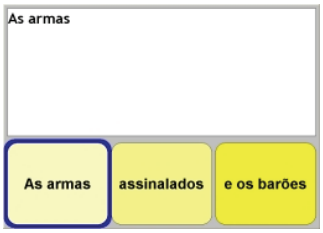


Figure 5: Virtual keyboard used to evaluate interaction with one continuous control signal to move mouse cursor in one-axis and one binary control signal to make key selection.

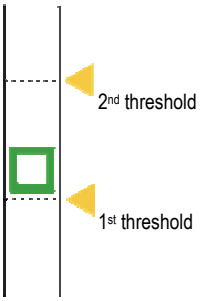


Figure 6: Representation of the technique used to combine two different control signals. By moving the bio-signal amplitude (the square) between the two thresholds, the user will move one object in one direction. When overcoming the 2nd threshold, the user makes a selection.

d) Two continuous control signals

Then user was asked to use two continuous control signals to navigate through rows and columns, in a keyboard as shown in Figure 7. Two different brainfingers (control signals generated by the studied user interface) were used. The source of these signals were muscle potencial generated by opening jaw and one brainfinger potencial (Junker, 1995) generated by subtle forehead movements.

Control was based on these two control signals: the first (described in Figure 6) to control x-axis, and the second to control y-axis.

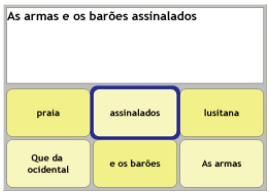


Figure 7: Virtual keyboard used to evaluate interaction using two continuous control signals to move mouse cursor in two-axis and one binary control signal to make key selection.

5 RESULTS

The user was able to control the UI using different selection methods. Qualitative and quantitative data were analysed, giving together a more complete evaluation of the results (for more details, please refer to (Londral, 2007).

The main problems related to control signals were low SNR, involuntary generated control signals and delay in generating the control signal. The latter was due to the difficulty in raising and lowering the amplitude of the control signals. After some minutes of training, the involuntary impulses were almost suppressed. When writing using a scanning access

method, and after five minutes training, user was able to do 3,18 key selections per minute. Considering that user is able to do 5,16 key selections per minute with his usual UI (a pressing switch) and that this result was reached in just one short session of training, it is expected that this performance will improve with training.

The user could also generate one continuous control signal using it to move the mouse cursor in one axis successfully. Resorting to a control signal that combines continuous and binary features, as described in Section 4.1-c, the results obtained were 4,36 key selections per minute, thus improving the performance attained with one binary control signal. However, it is important to note that this selection method was tested just with a small selection set (smaller than the one used for the previous result).

When testing the use of two continuous control signals, in order to control the mouse cursor in two axis (as described in Section 4.1-d), the performance was only 1,71 key selections per minute. This selection method was difficult for the user, especially in managing the control of the different thresholds. Therefore, more training is necessary to validate this technique.

From this case study, it is clear that the user was able to generate various types of control signals that could provide more *flexibility* to a UI, thus making it more adaptable to the user progressive conditions.

5.1 Blended Control Signals

Traditionally, control interfaces generate binary control signals (used to control scanning methods) or continuous control signals in 2-axis (used to control direct selection). Based on the various types of electrophysiological signals that the individual in this study could generate, a new class of control signals is proposed – *blended control signals* - that combine in a single signal discrete and continuous features. Based on these signals, different access methods can be designed. Beyond traditional selection methods, these signals can potentially fill the *gap* between scanning and direct access methods, as discussed in Section 2. In fact, the interaction described in Section 4.1-c) is neither direct nor scanning.

From this study was demonstrated that users may have potencial to generate control signals with more information than just for a binary control, though not enough to direct selection.

In progressive conditions, users experiment different needs and abilities along different stages. The more information the user interface can collect from users' abilities, the faster may be the access to AT systems.

The use of blended control signals, based on user's electrophysiological signals, allows a better adaptation to neurodegenerative conditions, broadening the possibilities of ways of interaction and enabling persons with severe neurodegenerative disorders to interact more efficiently with AT systems.

6 CONCLUSIONS

In this paper a case study demonstrating the use of electrophysiological control signals by a young man with ALS was presented.

The user was able to "upgrade" the control signals by progressive steps. Starting by a binary control signal using a scanning method, he was progressively able to generate continuous control signals, as well as combinations of these – *blended control signals*. The case study here presented clearly shows that other selection methods should be sought taking advantage of the control signals that this kind of users may be able to generate, in a sense richer than binary signals, although poorer than a continuous signal.

This kind of signals may provide more flexible and efficient ways of interaction with AT systems, if multimodal selection methods are designed. Moreover, resorting to blended control signals, AT systems may become more user friendly and adaptable, reducing the rate of AT abandonment, especially among people with neurodegenerative conditions.

REFERENCES

Allanson, J., 2002. *Electrophysiologically interactive computer systems*. Computer, IEEE Computer Society Press; 35(3): 60-65.

Azevedo, L., 2006. A Model Based Approach to Provide Augmentative Mobility to Severely Disabled Children through Assistive Technology. PhD thesis. Universidad del Pais Basco.

Cook, A., Hussey, S., 2001. *Assistive Technologies: Principles and Practice,* Mosby. 2nd Edition.

Felzer, T., Nordmann, R., 2006. *Speeding up Hands-free Text Entry*. Online-Proceedings of 3rd Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT), Fitzwilliam College, University of Cambridge. From: http://rehab-www.eng.cam.ac.uk/cwuaat/proceedings06.html. May 2007.

Glennen, S., DeCoste, D., 1996. *The Handbook of Augmentative and Alternative Communication*. Singular Press.

Iacono, T., 1992. *A Step Back from Small 'N' Design to View the Total Picture*. In: Gardner-Bonneau, D.J., Methodological Issues in Research in Augmentative and Alternative Communication. Proceedings of the Second ISAAC Research Symposium in Augmentative and Alternative Communication: 12-16.

Junker, A., 1995. United States Patent, 5,474,082.

Londral, A., 2007. *Tecnologias de Apoio para pessoas com Esclerose Lateral Amiotrófica: estudo de interfaces baseadas em sinais electrofisiológicos*. Master thesis. Universidade Católica Portuguesa.

Stevens, R.D., Edwards, A.D.N., 1996. *An approach to the evaluation of assistive technology*. Proceedings of Assets '96; ACM: 64–71.

United States Congress, 1998. The Technology-Related Assistance for Individuals with Disabilities Act of 1998. PL 100-407.

Wills, S. A. e MacKay, D. J. C., 2006. DASHER—An Efficient Writing System for Brain–Computer Interfaces? IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol. 14 (2).

Wolpaw, J.R., Birmauer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M., 2002. *Brain-Computer Interfaces for communication and control*. Clinical Neurophysiology; 113: 767-791.

A VISUAL INTERFACE BASED ON THE MVC PARADIGM TO LOCATE PEOPLE

Francisco Javier Ferrández Pastor, Juan Manuel García Chamizo and Rafael Jesús Valdivieso Sarabia
Departamento Tecnología Informática y Computación, Universidad de Alicante
Carretera San Vicente del Raspeig s/n, Alicante, Spain
fjferran@dtic.ua.es, juanma@dtic.ua.es, rvaldivieso@dtic.ua.es

Keywords: Elderly care services, wearable devices, indoor location, multi-frequency techniques, model-view-controller.

Abstract: Location and tracking of mobiles, i.e. people, vehicles and systems, where global positioning systems are unable need the use of ad-hoc local techniques. There is a vast field of applications that include children and elderly cares, location of specialists in hospitals and industry, security systems, etc. Our approach proposes supervisory software and radio frequency data acquisition system to offer location and tracking services. In this paper, we present a graphical interface for simulating, monitoring and managing what each scholar does and where he is. The proposal is based on the model-view-controller paradigm and a Java implementation has been developed.

1 INTRODUCTION

The developed societies are experiencing a significant ageing. The Statistical Office of the European Communities, EUROSTAT, states that 65th and older supposed 17 % of total population in the European Union in 2005. Estimations are made that 30 % will be reached at 2050 (EUROSTAT, 2006). This situation leaves us to contribute by mean of technological proposals, to reduce the cost of the services.

Our government is paying special attention to dependent persons in the last years. They supposed 9 % of the spanish population in 1999 (INE, 1999). The recent approval of the "Law of Promotion of the Personal Autonomy and Attention to Persons in Situation of Dependence" gives us the mark to develop services for disabled persons. Those services could require sanitary and welfare qualified personnel or technological support, too.

Children up to 9 years were 9.39 % in 2002 and 9.85 % in 2007, showing a light increment. The security of children is being one of the most relevant objectives of his parents and the authorities. As a consequence, children are using mobile communications early.

Location systems can also be used in hospitals, commercial or leisure centres, fairgrounds, as well as to locate doctors in hospitals, workers in mines, watchmen and so on.

The main objectives of this work are provide a visual interface for simulating, monitoring and managing computer utilities oriented to locate and track people and mobile targets. That tool should be friendly to satisfy monitoring requirement, precise to support experimentation and powerful to configure utilities in an easy way.

Following, second section describes generalities of the location system, and an overview of more known outdoor and indoor location systems is made. In section three, we explain a module of utilities oriented to scholars and screen captures are showed to suggest the reader its potentialities. Conclusion section indeed the main goals and related future works.

2 LOCATION SYSTEM

The location system takes charge verifying the position of each one of the persons to monitoring. There develop services derived from this location depending on the requirements of the place where has been installed.

The location system can be used in centres where the position of the persons should be known in each time. This information could be useful to increase the security of the hospitalized persons. It could be used to improve response time of critical services like hospital urgencies, also.

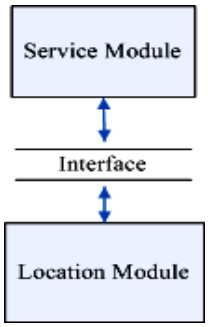


Figure 1: Basic scheme of the location system.

The structure of the location system is determined by the functional organization that can be observed in figure 1.

The location module takes charge to verify the position of each one of persons. The other one, Service Module, takes charge providing the services derived from position. The independence between modules allows us replace the location module by future versions.

The location module takes charge determining the real position of the transmitter respect of the receiving antenna. This way, knowing the location of the antenna, we manage the position of every person. We must choose the place where persons are going to take the transmitter. This is important, because it could determine the success or failure of the system in some environments. Nowadays there are devices that are placed in the wrist or in the neck like a hanger. Persons are conscious that they are carrying it. To avoid this situation, we think that the transmitter should be placed inside the footwear, inside the heel. That allows the users don't be conscious about the transmitter. To place it in the shoe, the system must have autonomy. It will not be possible recharge it during his useful life. Therefore, in the choice of location system, the energetic consumption and the size will be determinant.

We are going to analyze the different location technologies. We can classify them under two big groups: outdoor location system and indoor location system.

The maximum exponent of outdoor location technologies is Global Position System (Nord et al., 1998). It coexists with the incomplete system GLONASS (Langley, 1997), and it will coexist with GALILEAN (European Commission, 1999). All of these global positioning systems have very small coverage inside buildings so in this context the location it is not possible.

Inside indoor technologies are based on wireless technologies of communication. Using some features about these wireless technologies we can determine position of the transmitter. First we are going to emphasize a system based on the wireless technology bluetooth (Bluetooth SIG, 2001), named Bluetooth Indoor Positioning System (Dark and Delmastro, 2001). The core of the system is a network of bluetooth access points, distributed inside the building every 10 meters. The system is coordinated by a central computer. This system is appropriated for small buildings due to the limited coverage of every node. We need a great number of nodes. The positioning is realized by means of sign triangulation. It is necessary synchronize to all the access points.

RADAR (Bahl and Padmanabhan, 2000) is based on the standard IEEE 802.11 Wireless LAN (IEEE, 1999). It combines empirical measures with models of the sign spread, to determine the location user. This system provides more precise results than the previous though reverberating in a high cost.

By means of the technology named Transmission Error Approximation (Xiao et al., 2006), the distance is determined between two zigbee devices. They measure up and analyze statistically information of the packets that have not arrived to their destination. The transmitter sends different packets each one with a size and different energetic level. Therefore, this technology is not adapted for systems that need great autonomy.

Leaving aside the systems based on wireless technology, we meet a system named GETA Sandals. (Yeh et al., 2007). It is composed by two sandals equipped with force, ultrasonic, orientation, RFID sensors and accelerometers. It allows verify the position from the initial point, adding placement vectors that are formed while you are walking. This system needs few infrastructures, but presents some difficulties. The travelling error is accumulating, so it is necessary to put some control points.

Each of the previous technologies presents a few advantages and disadvantages for our objective. We are developing an alternative indoor location technology based on the drift that suffers the multi-frequency electromagnetic field when cross a well-know medium, (Ferrández, 2007), (Pujol et al, 2005), (Ferrández et al, 2007). It works emitting an electromagnetic multifrecuencial signals, and then we measure the delay that is produced between the frequencies on having crossed the way. By means of this technique we can determine the distance using only a receiving aerial, avoiding the triangulation. In addition the transmitter only must emit a pulse periodically; therefore the consumption will be low. The delay that suffers each of the frequencies is in

the range of nanoseconds. The cost of the equipment capable of measuring it is high.

The services module is the application that simulates the services derived from the position, manages the persons to locating, and monitors the position throughout the time. This application is going to be used by the final user.

We propose the development of a common base application useful to all the areas where the system could be implanted. The services module must show a plane of the building and each person in his current position. Thinking about geriatric centres or colleges, we will want that the system warns us if someone goes in or goes out of a certain zone. Also we will wish that the system learns the behaviour of each one. If we think about a hospitable environment, a possible service would warn automatically the more nearby doctors to the urgency.

3 SERVICES MODULE

The visualization environment provides mechanisms to manage, to monitor and simulate the services derived from the children's location inside a college. In figure 2 we can observe it when it is launched.

The implemented services are: S1, to know every time where each person in the building is; S2, to know where a person is at the present time; S3, to know the information of the persons to locating; S4,

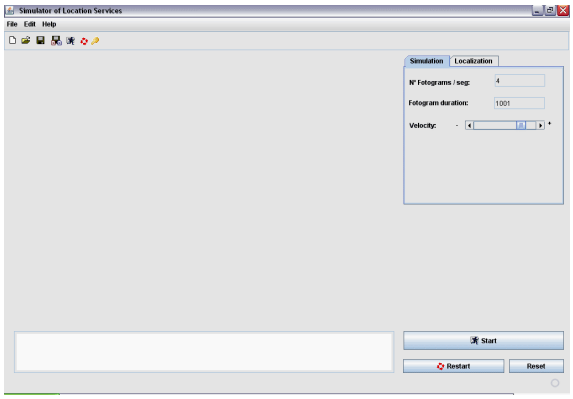


Figure 2: Environment aspect when is launched.

to know where the persons have been throughout the time. Saving and analyzing these positions data we will obtain standards of behaviour; S5, warns us about an accident; S6, the system warns us when someone goes in a room, which is in secure way. The environment provides the services previously described by means of graphical interface. They allow using any service. In the figure 3 we can identify the main widgets.

The principal graphical visualization, F1, takes charge representing the facilities of the building and the position of each one of the persons in the current time. It allows to select graphically form the persons and to put them along the plane, to locate them in another stay, even while the simulation is realized. It allows select the stays to establish them in secure way or in normal way.

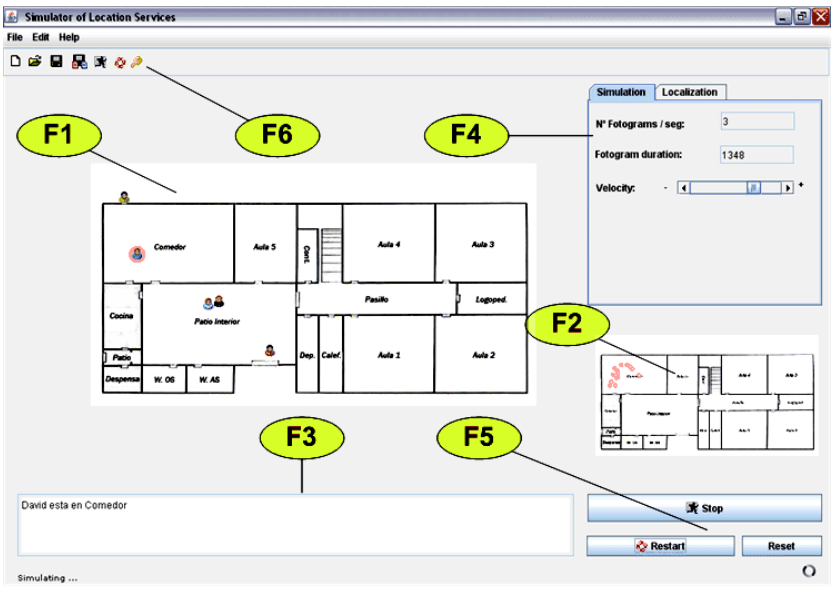


Figure 3: The window contains the main widgets to control the application.

The secondary graphical visualization, F2, represent the plane from the building a smaller scale, reflecting graphically the positions of the selected person.

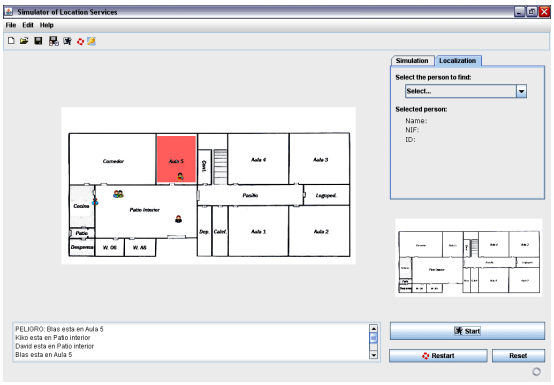


Figure 4: State of the environment when a person is inside a secure room.

The visualization in way text, F3, shows the incidents and the position of all the persons. When a person has been selected, only the information about the position of this person is shown.

It is possible to manage the frequency of the simulation. The frequency is specified in milliseconds. In every period the new position of the persons is calculated and the changes produced are represented in the environment.

There are a set of controls widgets, F5. They allow to start, stop or to restart the system.

Activating the sure way, F6, the system will warn us when someone goes in one secure room. In the figure 4, we can see one image of the environment informing us about an incident. This incident shows a person inside the secure room. The room is shaded and a text message is showed.

3.1 Structure

The environment is structured to facilitate the maintenance, the reutilization of the source code and the detection of mistakes.

The modular division has been realized to facilitate the incorporation of new services. We have used the design model Model View Controller, MVC, (Sommerville, 2004).

The environment is structured in three layers: graphical interface of services, logic control and installation data. It is possible to observe in the figure 5 the correspondence between the pattern MVC, the environment layers and the environment elements. The graphical interface of services layer corresponds to the view layer of the model MVC. It defines the graphical user interface. The controller corresponds to the logic control layer. It is the logic of the application. It takes charge managing the time, the events of representation, the visualization of the objects of the scene, and the beginning and the

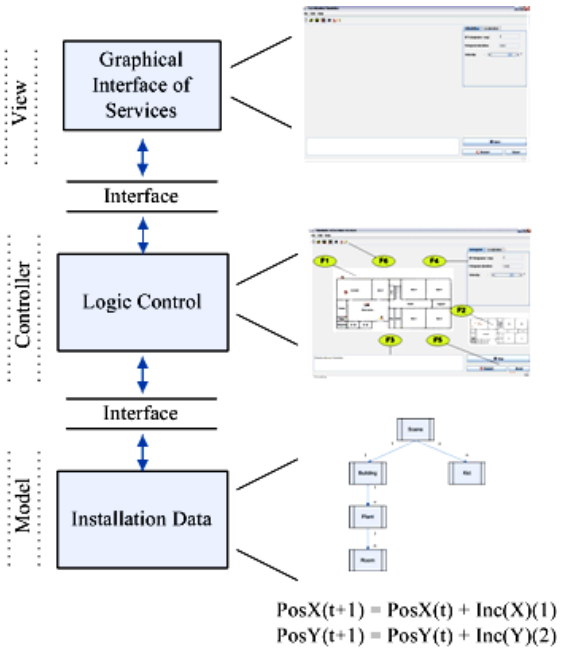


Figure 5: Correspondence of the environment layers and the elements with the model MVC.

end of the simulation. It uses the events of user. The user can interact asynchronously with the control widgets: F1, F2, F3, F4, F5 and F6. The model is represented by installation data layer. It takes charge storing the necessary information to represent the facilities of the college and children, and defines events of representation for each one. The information is formed by a hierarchy of entities grouped under the entity Scene. Scene contains a list of entities Kid. They represent the children who are inside the college. The entity Scene also includes the entity Building that represents the building and contains a set of entities Plant. These entities represent the floors of the building. And each of these contains a list of entities Room. The entity Room represents the rooms of the floor. This hierarchy can be observed in the entity relationship graph showed in figure 6.

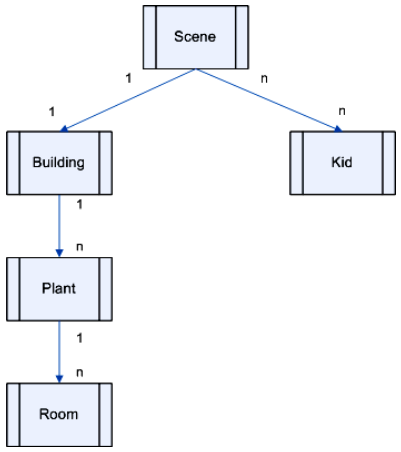


Figure 6: Entity relationship diagram of the scene.

The representation events are executed when it is necessary to update the person's position. Later, the general state of the scene is determined, and the changes are represented in all the views. In the prototype, we simulate the movement of the persons, We will establish the starting point of each person and later when beginning the simulation, each one of these people will move randomly. The random movement has been implemented from the present position, PosX (t). To this position a random displacement is added to him, Inc. (X), obtaining therefore the new position, PosX (t+1), for the following moment of time. This one calculation will be made for each one of the coordinates. In equations 1 and 2 we can observe the calculation of coordinate X and Y, respectively.

$$PosX(t+1) = PosX(t) + Inc(X) \quad (1)$$
$$PosY(t+1) = PosY(t) + Inc(Y) \quad (2)$$

3.2 Implementation

The programming language is Java because is a multiplatform, object-oriented and facilitates integration in web platforms.

The environment is organized in a set of layers. Every layer is implemented by means of a set of classes. In the figure 7 it is possible to observe the corresponding class diagram. The graphical interface of services layer is composed by 4 classes. GraphicInterface takes charge managing all the elements of the graphical user interface. PictureArea takes charge realizing the representation in the graphical principal visualization F1. MiniPictureArea is a subclass of the previous one, and it overwrites his visualization method, because in graphical secondary visualization F2, we show different information from the parent class. ProgressCircle takes charge representing a circle animation. The animation is visible while the simulation is on. The logic control layer is formed by unique class. It is called application and contains the simulation engine, the services that it offers to the user, and the communication with the graphical interface of services layer and installation data layer. The base of the simulation engine is implemented by means of a thread clock. This clock is configured in milliseconds. In each event the new position of each person is obtained, the model is modified and the new scene is represented.

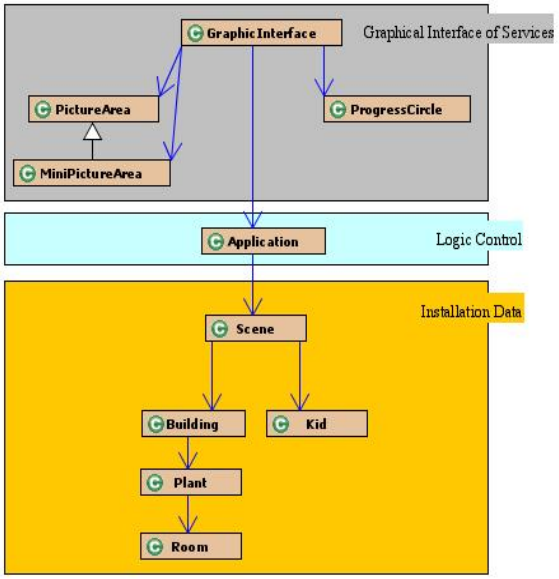


Figure 7: Class diagram for each one of the layers that forms the service module.

The installation dates layer is formed by 5 classes. They represent the scene to simulate and

contain the information defined in the entity relationship diagram showed in figure 6. The representation of the scene is in charge to update the state of all the objects that take part, determining the interactions to each other. In the first place the plane of the plant of the building is drawn. Later the safe rooms are marked by means of a shade of yellow colour. If a child has entered some secure room, it will be marked with a red shade. If the user has selected some child, then a red circle will be drawn in the child position, and finally every child are drawn in its respective position.

4 CONCLUSIONS

In this paper, we have presented an interface to assist a location and tracking system that solves the problem when global positioning systems are unable. We have shown the potentialities of these services to satisfy the necessities in schools, hospitals, industries and old people's centres and we have introduced a simulation tool that serves as user interface and to configure the system.

We consider a software implementation in Java, because it offers a platform independent and it has many networking capabilities to build distributed systems, for example to remote data acquisition from sensors.

As future works, we will work to evaluate some of the main indoor location system to integrate it and add new features to the module services.

ACKNOWLEDGEMENTS

This work has been partially supported by spanish government as I+D project: Desarrollo de herramientas CAD/CAM para el prototipado virtual en el sector calzado. DPI2005-09215-C02-01.

REFERENCES

Bahl, P. and Padmanabhan, V., 2000. *RADAR: An in-building RF-based user location and tracking system*. *Proceedings of the IEEE Infocom 2000*, March 2000, Tel-Aviv, Israel. 775–784.

Bruno, R. and Delmastro, F., 2003. *Design and Analysis of a Bluetooth-Based Indoor Localization System*. In: Springer Berlin / Heidelberg eds. PersonalWireless Communications. Berlín: Springer, 711-725.

European Comision, 1999. *Galileo Involving Europe in a New Generation of Satellite Navigation Services*. Brussels: European Comision.

European Comision Eurostat, 2006. *EU25 population aged 65 and over expected to double between 1995 and 2050*. Luxembourg: European Comision Eurostat.

Ferrández, F.J., 2007. *Deriva frecuencial de la transmisión electromagnética por efecto del medio*. Thesis (PhD). Alicante University.

Ferrández, F.J., Valdivieso, R.J and Albert M..M. 2007. FootLose, zapatos para la localización de personas. FUNDEUN, University of Alicante.

IEEE, 1999. *IEEE Standard for Wireless LAN. Medium Access Control and Physical Layer Specification*. P802.11. Piscataway:IEEE.

IEEE Standard Association, 2001. *Bluetooth SIG. Specification of the Bluetooth System, Volume 1–Core. Version 1.1*, New York:IEEE.

Instituto nacional de estadística de España, 1999. *Encuesta sobre Discapacidades Deficiencias y Estado de Salud.* 1999. Madrid, Spain.

Langley, R. B. 1997. *GLONASS: Review and update*. GPS World, July, Vol. 8, No. 7, 46-51.

Ministerio de Trabajo y Asuntos Social, 2005. *Atención a las personas en situación de dependencia en España, libro blanco*. Madrid: Instituto de mayores y servicios sociales.

Nord, G., Jabon, D., Nord, J., 1998. *The global positioning system and the implicit function theorem*. In: R. Sincovec, ed. SIAM Review, v.40 n.3. 692-696.

Sommerville, I. 2004. *Software Engineering*, 7th ed. Boston: Addison-Wesley.

Pujol, F.A., Ferrández, F.J., García Chamizo, J.M., Fujol, M. and Rizo, R. 2005. *Numerical analysis in position location: a frequency-domain model.* Kybernetes: The intenational Journal of Systems & Cybernetics. Vol. 34, 704-710.

Xiao, W., Sun, Y., Liu, Y., Yang, Q., 2006. *TEA: Transmission Error Approximation for Distance Estimation between Two Zigbee Devices*. International Workshop on Networking, Architecture, and Storages, 2006, Shenyang, CHINA, 15-22.

Yeh, S., Chang , K., Wu, C., Chu, H., Hsu, J.Y., 2007. *GETA sandals: a footstep location tracking system*. In: Springer London eds. Personal and Ubiquitous Computing. London: Springer, 451-463.

HEALTHCARE IN CONTINUUM FOR AN AGEING POPULATION

National Self Monitoring or Remote Offshore Monitoring for Australia?

Heather Grain, Kerin Robinson, Belinda Torney
HIM Program, School of Public Health, Faculty of Health Sciences, La Ttrobe University, Bundoora, Vic 3058, Australia
H.Grain@latrobe.edu.au; K.Robinson@latrobe.edu.au; belinda.torney@latrobe.edu.au

Bardo Fraunholz, Chandana Unnithan
School of Information Systems, Faculty of Business and Law, Deakin University, Burwood, Vic 3121, Australia
bardo.fraunholz@deakin.edu.au; chandana.unnithan@deakin.edu.au

Keywords: Ageing in Australia, Electronic Health Records, Privacy, Offshore Resourcing.

Abstract: Australia is a country, similar to other developed nations, confronting an ageing population with complex demographics. Ensuring continued healthcare for the ageing, while providing sufficient support for the already aged population requiring assistance, is at the forefront of the national agenda. Varied initiatives are with foci to leverage the advantages of ICTs leading to e-Health provisioning and assisted technologies. While these initiatives increasingly put budgetary constraints on local and federal governments, there is also a case for offshore resourcing of non-critical health services, to support, streamline and enhance the continuum of care, as the nation faces acute shortages of medical practitioners and nurses. However, privacy and confidentiality concerns in this context are a significant issue in Australia. In this paper, we take the position that if the National and state electronic health records system initiatives, are fully implemented, offshore resourcing can be a feasible complementary option resulting in a win-win situation of cutting costs and enabling the continuum of healthcare.

1 INTRODUCTION

Ageing of populations is a world-wide phenomenon (Ozanne et al., 1997). However, in the past few decades, the central characteristic of ageing in societies worldwide has changed from high birth and death rates, to lower birth rates and increased life expectancies (Rowland, 1991). In Australia, failing fertility rates combined with the baby boomer generation moving into old age groups have contributed to the irrefutable demographic change (ALGA, 2005).

The proportion of people aged over 65 years, which is currently 13 percent (2.5 million), is expected to grow to one quarter of the population by 2051. While the proportion of people over 85 years is expected to grow from the current 1.4 percent to 6 percent during the same period, the people within the workforce age of 15-64 years is expected to fall from the current 67 percent to 59 percent (ABS-01, 2007). As a reflection of fertility, mortality and migration, population ageing in states and territories shows varied trends (ABS-02, 2007). A significant dimension of the ageing population is the multiplicity of needs, interests and backgrounds. A wide range of inter-related factors including gender, location, socio-economic status, general health, culture and education have influenced the ageing process of individuals.

As Ozanne et al. (1997) recounted, migration resulting in the multi-ethnic character of the population seems to imply a need for differential arrangements in the public provisioning of health services as the ethnic aged may not share the attitudes of the mainstream groups. Conversely, the baby boomer generation is expected to age with different aspirations and expectations and on average greater financial means than previous generations (Australian Government, 2005). In addition, this generation is growing with transitional evolutions in technology. Their expectations of

independent living, for a longer period, are much higher compared to those of previous generations.

In a recent research study, "The Economic Implications of an Ageing Australia", the Productivity Commission (2005) concluded that the delivery of human services, which represents 49 percent of local government expenditure, is forecasting the main demands from healthcare provisioning for ageing and aged care. Further, it reiterated that local councils are more likely to face budgetary pressures from population ageing than from traditional activities such as infrastructure provisioning.

The development of the "Australian Government National Strategy of an Ageing Australia" has provided a framework for responding to the opportunities and challenges of population ageing (Australian Government, 2005). It implied that population ageing affects more than aged care services and that an effective response requires a holistic approach including local governments. While Australia operates various initiatives for specific diseases and risks, the Australian Government (2005) has called for a more comprehensive approach to supporting and promoting health throughout life, which will require effective action across the continuum of care. Prevention and management of ill health are therefore seen as complementary strategies.

How can a continuum of care be facilitated to cater to the privacy conscious, financially sufficient, technology savvy generation with an ethnic mixture of migrant population? Are the current practices or initiatives ongoing, in alignment with their expectations of self-monitored, independent living for a longer term? To explore answers to these questions, in the next section, we consider the composite of Australian population: both the ageing and the aged, and ongoing initiatives.

2 AUSTRALIA IN PERSPECTIVE

One of the key expectations of the ageing population worldwide is a non-intrusive, privacy facilitating approach to health care. The generation that is ageing today has grown with digital technologies and is able to self-monitor their health to a significant extent. In an ideal situation, healthy ageing is preferable to after-care, which has been recognised as a major pressure on national budgets. We take the example of USA-based research to demonstrate the point.

Hayes et al. (2003) reported that over 20 percent of the USA population in the 85-plus age group were found to suffer from malnutrition and medication non-compliance as they do not receive appropriate interventional treatments by medical practitioners. This segment had limited capacity for independent living, with the result that they required continuous monitoring and daily care. This realisation spurred the piloting of remote monitoring of the activity of people in their homes for detecting acute events such as falls, using unobtrusive techniques (Glascok et al., 2000; Ohta et al., 2003).

The Point-Of-Care Engineering Laboratory at OHSU is one among the pioneers in developing approaches and technologies that allow early detection of reduced physical and cognitive function that leads to decreased independence (Hayes et al., 2003). They identified three key factors that facilitated this activity: the technologies must be unobtrusive, flexible and adaptable; and provide complete privacy. They argued that if people are aware of technologies, behaviours change and self-monitoring is un-reliable. Installing un-obtrusive and inexpensive sensors is perhaps the answer. Second, technologies needed for health monitoring are probably not those people would want in their homes all the time. Therefore, wireless devices and open standards for device communication are essential to simplify the placement of technologies on a needs basis and to meet dynamic health care needs. Finally, the use of encryption and authentication techniques including pseudonomisation, as well as judicious selection of what information is actually transferred between devices on the network, are recommended.

Advances in wireless networking, ubiquitous computing and unobtrusive technologies are now providing opportunities for facilitating healthy ageing and aged care, both of which are in the portfolio of health care provisioning for nations.

2.1 Initiatives

In Australia, two distinct issues confront governments in healthcare provisioning: providing for healthy ageing, to ensure a nation of self-independent individuals; and addressing the needs of the already aged proportion which needs assistance for living. We examined initiatives and frameworks that are ongoing for both categories.

There is a significant move towards e-Health recognising the facts that the society is increasingly technology savvy and that people would like to have the option of self-monitoring their health. In late

1999, the Australian National Health Information Management Advisory Council (NHIMAC) released a "Health Information Action Plan for Australia", which constituted the use of online technologies within the health sector and laid out national projects (Ride, 2007). A key recommendation was to develop a national framework for the use of e-Health records to improve the efficiency, safety and quality of care within the requisite privacy legislations.

The initiative of HealthConnect included the establishment of a national framework for a system of electronic health records – which involves the electronic collection, storage and exchange of consumer health information via secure networks and within privacy safeguards. The network, with consumer consent, allows electronic exchange of clinical information between health care providers. The information regarding consumer health was to take the form of standardised 'event summaries', extracted from health care provider electronic records for consultation, including current and historic information such as results of pathology and diagnostic tests, hospital discharge summaries, chronic illness monitoring, current medications, allergies, immunisation information and principal diagnosis (Ride, 2007).

Rather than replacing existing legacy systems at national, state and territory levels, HealthConnect proposes a composite of different layers: a records layer which consists of the regional storage repositories; a user layer which contains the computer systems software that will interface with HealthConnect to allow providers either to view/review records or add new event summaries to records; and a national coordination layer that links all regional storage repositories and provides the link between these repositories and user/source systems, allowing people to use their record throughout Australia (HealthConnect, 2007a) .

However, the challenges for the system are still numerous. For instance, the states and territories are not alike, and there are major differences in geographic size, population density, investment capacity, information management infrastructure and health care issues and resources (HealthConnect, 2007b). As a result, each region is at a different stage of implementation development (Ride, 2007).

HealthConnect works briefly as follows: peoples' information is collected at point-of-care. The shared storage facility helps health care providers to access information, with consumer consent. Conversely, an individual may wish to check progress against key self-management observations such as blood pressure, weight or blood glucose levels. According to Ride (2007) who provides the latest status, this network ensures robust privacy and security standards. The expectation is that by 2008, Australia will be well advanced in achieving the goal of electronic connectivity between all major health institutions and health care providers (HealthConnect, 2007b). From a future perspective, national implementation within 18 months (as of 2007) is expected in some areas such as e-prescriptions, e-referrals and hospital discharge summaries.

An Australian consortium is developing a possible architecture known as OpenEHR (OpenEHR, 2007). Based on openEHR release 0.9, a scalable, secure, shared e-Health record to meet national standards is being implemented using a combination of XML, Web Services, J2EE, Relational database, LDAP and PKI. Via HL7, it also supports interfaces with external systems and smart cards. The totally web-driven user interface works with all popular browsers. It currently supports hospital doctors, general practitioners, pathologists, endocrinologists, ophthalmologists, dieticians, diabetes educators and podiatrists. Emergency medication, pharmacies and community nursing are being added (OpenEHR, 2007).

In 2002, Standards Australia published the AS5017-2002- Health Care Client Identification to provide a basis for improved association of clients and their data between organisations. There are currently two dominant sets of messaging standards in the Australian health sector: UN/EDIFACT for financial applications; and Health Level 7 (HL7) for more clinically-related applications. Other standards such as DICOM (Digital Imaging and Communications in Medicine) are applicable to discrete applications such as diagnostic imaging. Standards Australia International is developing a Message Usage Handbook that provides recommended applications of the messaging standards. Further development of this message usage model is anticipated in the short to medium term (Ride, 2007).

In July 2004 the National E-Health Transition Authority (NEHTA), a not-for-profit company, was established by the Australian, State and Territory governments to develop better ways of electronically collecting and securely exchanging health information. Its mission is to set the standard, specification and infrastructure requirements for secure, interoperable, electronic health information systems (NEHTA, 2007). The Australian initiatives for monitoring health in the ageing are being built on

the existing principles of privacy, legislation and standards.

Conversely, there is increased pressure on the national budget to cater to the 85-plus or significantly non-independent aged population. Lifestyles have catered to longevity and perhaps technology-assisted living. Soar et al. (2007) reported the current status on approaches to reduce avoidable hospital admissions through information technology. The aged Australian prefers home care referrals, a less expensive alternative to institutional care. A workshop at the 2005 Health Informatics Conference identified a lack of reliable identification of candidates for hospital avoidance as a major barrier.

Subsequently, the Advanced Community Care Association (ACCA) was formed to provide a single point of referral to community service organisations. Further, "Nexus eCare" developed by Nexus Online Pty Ltd, provided a proof-of-concept web-based, community care management system which identified candidates, mapped services to patients, automated communication between hospitals and community service providers. Initially, the system uses a Rapid Assessment tool to identify avoidable patients. It incorporates an "intelligent filtering agent" which continuously monitors the digital data flow. The final assessment of this approach reveals that pressure on hospitals, emergency departments and budgets can be significantly relieved (Soar et al., 2007).

Philipson and Roberts (2007) reported on the impact of technology on aged and assisted living in Australia. This research recognised that the usage of digital technologies by the aged will become increasingly an important issue in future years. The authors point to a number of proactive computing applications that are being developed which will assist ageing persons to live longer in their home environments. Assistive technologies are wide-ranging, from radio/ultrasound/remote control appliances, alternate keyboards, voice input devices, phone amplifiers, etc.

Wireless sensors, for example, can be used to gather behavioural and biological data, to be input into computer applications (Philipson and Roberts 2007). Conversely, virtual uninterrupted communication possibilities as the user moves from their homes to cars or external places are being envisaged, with Telstra and other carriers in Australia contemplating a next-generation network (NGN). The use of BANs (Body Area Networks) can be useful in assisting home monitoring of paraplegics or compensating for deficits of functioning caused by dementia. BAN, a base technology for permanent monitoring and logging of vital signs, is a proven method of supervising the health status of patients suffering from chronic diseases, such as diabetes and asthma (BAN, 2007).

Hovenga et al. (2007) have described comprehensive and recent research developments in the area of ageing/aged care using OpenEHR. They proposed an archetype management framework to facilitate the development of future information systems and optimise electronic health records within the aged care sector. According to them, Australia is leading in the field of developing Electronic Health Records using openEHR archetypes. These archetypes describe rich information structures by indicating how the information is to be expressed; what is optional and mandatory; what is a sensible value for each data element; and other rules (Hovenga et al., 2007:4). These archetypes have the potential to improve aged care in many ways such as standardising clinical content and enabling the data to be interchangeable; empowering residents by enabling them to switch providers easily without the need for multiple examinations; improving provider access to relevant resident information; providing necessary flexibility to reflect resident care preferences; and enabling care providers to access best practice information as part of daily workflow and decision making processes at the point-of-care (Hovenga et al., 2007:4)

Both the aged and the ageing would benefit from a nationwide semantic interoperability, requiring the national adoption of a key set of standards. Standard openEHR archetypes include the adoption of a standard terminology and set of data types, which best fit with the openEHR information model but can be used, to a variable extent, to enable communication between systems with different information models. Currently, an international team lead by Australian experts is engaged in identifying a common standard set of health data types and encouraging their adoption into international standards.

The current clinical information systems tend to be vendor specific, not adopting standard data models, due to the lack of agreed standards. Hovenga et al. (2007) recommend the adoption of standard structured messages that are compliant with messaging standards such as developed by Health Level 7 (HL7) and its international affiliates. Further, Standards Australia has developed a number of HL7 standard implementation guidelines for this purpose (Standards, 2007).

Having explored the initiatives for aging and the aged for future, we now look at a critical factor in the healthcare provisioning - privacy.

2.2 Privacy

Moor (1997) suggested that privacy is felt when a person is protected from intrusion, interference and information access by others. In most western societies, including Australia, this definition may describe privacy, but it is not a universally accepted concept. In many Asian societies shielding a person is not considered correct. In the context of Australia, which has a significant migrant population from these societies, it is regarded as acceptable only in rare circumstances to have the individual's privacy violated for general welfare.

For example, in Australia, RFID is used for non-invasive monitoring (Frost and Sullivan, 2005). However, does it concur with the needs of privacy that the ageing population expects today? Where individual privacy has to be respected, there are arguments for and against the use of RFID for monitoring or, for that matter, any forms of technology. ICTs pose a unique threat to personal privacy because of the type and quantity of personal information that can be collected, combined with the speed of transmission and length of time that the information can be held (Tavani, 2004:118). RFID monitoring intensifies ICT-related difficulties in protecting private information by offering the information collectors the benefits of ubiquity coupled with secrecy (Wiebell, 2005).

Mulligan et al. (2007:2) reported on research within the privacy regulations, detailing issues and solutions for custodians. They point out that there are two main ways in which data custodians handle the demands of privacy protection. The first is to seek informed consent and the second to respond by developing mechanisms for ensuring privacy i.e. the data are sufficiently de-identified and protected such that they cannot be linked back to the individual. However, Mulligan et al (2007) reiterate that this solution does not allow people to control the use of their data, nor to minimise the potential for individuals to be harmed.

One methodology (Mulligan et al. 2007:2-3) involves the separation of personal identifiers from clinical information and their separate encryption by the reporting clinician, and submission of these data to a "trusted third party" who allocates an identifier specific to paired data items and forwards personal identifiers and clinical data to separate repositories (Churches, 2003). As a result, the clinical data from disparate databases can only be linked by trusted third parties. For example, personal identifiers are not provided with specific disease registers. Mulligan et al (2007) reiterate that this method requires legal protection and financial support from government. They also point out that in Western Australia, the custodians of disease registries and health databases that contain personal identifiers and clinical information sign a Memorandum of Understanding authorising the third party (the Linkage Unit) to identify data concerning the same individual in different databases. The Linkage Unit allocates Unique Anonymous Identifiers for each individual.

In August 2007, a regulation occurred in the Federal Parliament of Australia, titled "1.1 Medicare Australia (Functions of CEO) Amendment Direction 2007 (No.2)". In plain terms, without contest or assessment of value for money, Medicare (the national health insurance program) can scope, develop, build and test the NEHTA Unique Healthcare Identifier program. The regulation authorises NEHTA to make a copy of the two key identity databases supported by Medicare Australia (the client and the provider databases) and use them to provide an identity service. Despite the prohibition in the Commonwealth Privacy Act (2000) of personal information being used for purposes other than for which it was collected by Government Agencies, it has been decided that information that was collected to enable Medicare benefits to be paid is to be used to operate the NEHTA UHI (AHIT, 2007).

The implications this has, for the trust the population will have in Medicare Australia to keep their private information private, are profound (AHIT, 2007). Some pertinent questions that the regulations bring forth are:

1. Where is the Privacy Impact Assessment that validates this approach?
2. Who will be responsible if there is a security breach and personal details are released and the individual is harmed?
3. How will the information be protected from unwanted disclosure or access?
4. What is the legal liability?

Medicare patient and provider databases are key sources of a healthcare identifier regime being introduced to support a shift to e-Health programs. Consequently, records belonging to 99 per cent of

Australians are contained in Medicare's Consumer Directory Maintenance System, considered to be the most up-to-date and accurate government repository of personal information. The law prevents the use of Medicare data for other purposes; however, the Human Services Minister has unlocked access via a legislative amendment tabled in Parliament on August 16 (AHIT, 2007).

The durability and applicability of current legislation relating to privacy is very much in focus for healthcare provisioning in the continuum of care for ageing Australia. Furthermore, the OpenEHR initiatives need to be implemented in conjunction with continued amendments in legislation and standards, to ensure privacy protection.

Now we look at another dimension of healthcare provisioning that is under consideration, i.e. offshoring.

3 OFFSHORE RESOURCING – IS THERE A CASE FOR AUSTRALIA?

In 2004, Curtin University of Technology initiated a project that examined the long-term feasibility of off-shoring to India. The considerations included health services such as radiology and diagnostics (CBS, 2004). The study is ongoing and results are yet to be published.

The ACCI (2005) recounts that in the transition to a globalised economy, offshoring is a viable option for businesses in Australia, to better manage costs and quality of services. In the health services arena, including medical, dental, nursing diagnostics and health services such as data entry, hospital administration and processes, offshoring is a viable alternative. For example, Indian medical practitioners are interpreting radiological scans for patients in the USA. The Phillippines is providing medical record transcription services to a number of developed nations. However, offshore resourcing may not be an option where physical presence is required for legal and/or practical reasons, such as health care regulations.

Conversely, Australia has a shortage of doctors and nurses, especially in regional areas (AusGov, 2007). The ABC News (2006) reported an acute shortage of hospital beds, as citizens spend at least 25 minutes in an emergency waiting room. The report says the top five reasons for medical admissions in public hospitals were: respiratory problems; cardiology and interventional cardiology; childbirth; renal dialysis; and neurology. While it may not be possible to offshore some of these services, non-critical procedures such as monitoring an asthmatic patient at home, could be done from overseas with relative anonymity.

For example, if effective pseudonomisation standards are finalised and therefore neither the person's identity nor the location is disclosed, we argue that there is potential safety for the aged to be monitored from overseas, given that the person is not within the limits of harm nor is of interest to the offshore partner other than those via their work commitments. Certainly, there is a case for feasible options in health services offshoring to enable continuity of health care.

4 OUTLOOK

Drawing conclusions from the above sections, it is evident that: if there is a functioning, fully implemented OpenEHR system that enables e-Health, supported by legislation and ongoing upgrades of standards that ensure privacy protection in Australia, this would enable healthcare in continuum for the ageing population.

Shareable EHR's such as those proposed in HealthConnect that are built on OpenEHR will, if fully implemented nation-wide, have the potential to address the requirements of a technology savvy, privacy conscious ageing population who expect their health services to enable independent long-term living. However, the key factor is that the system needs to be fully functional, implemented, and compliant with legislation and standards, nation-wide.

At the next level, we consider the budgetary pressure on national/local governments in the provisioning of health services as well as the shortage of doctors and in-patient facilities in hospitals, in Australia. Currently, while the debate on offshoring is still rampant, we argue that it can be a viable option based on the following.

Use of architectures such as those proposed by OpenEHR would enable non-critical processes to be separated from critical procedures that cannot be offshored. Subsequently, as described by Mulligan et al (2007), the separation of personal identifiers from this information and submission of relevant data to a trusted offshoring partner, via a dedicated portal, is easily possible. The offshore partner on the other end can allocate identifiers specific to paired data items and forward personal identifiers and clinical data to separate repositories. No doubt, the

issues of trust and reliability become significant in such cases.

A specialist offshore radiologist, for example, who is providing expert opinion based on radiology reports, does not need the patient's name. In another example, if an aged person is being monitored in their home environment by a medical practitioner overseas, only the details that are relevant to their health condition need to be revealed and other personal details can remain anonymous to the observer. An online portal can be used for a medical practitioner at both ends to input radiology reports, diagnostics and analysis sheets. A general practitioner at the Australian end could be the interface and connecting point through a shareable EHR.

Where the portal is interfaced with an EHR that meets the NEHTA Privacy Framework requirements it should provide legal certainty. The identities of people should not be disclosed to third parties – enabling privacy protection. The separation of selected non-critical procedures from the EHR to offshore providers should ensure that control remains within Australia, and can be updated seamlessly via an interface on the online portal. At that level, compliance with legislation and standards would also be met via a registered medical practitioner who would act as the interface.

Therefore, our position is to implement a shareable EHR, compliant with standards and legislation and then to resource services offshore, with anonimised records, to improve service turnaround, relieving the pressure on government budgets and the skills shortage, while enabling healthcare in continuum for a healthier ageing Australia.

REFERENCES

ABC News (2006). Hospital bed shortage worsens, ABC Online, Retrieved September 10, 2007, from http://www.abc.net.au/news/stories/2006/07/03/1677583.htm

ABS-01 (2007). Population Projections *Australia 2002-2101 Cat.No.3222.0*, Australia.

ABS-02 (2007). Population Projections – Statistical Local Areas *1999-2019 Cat No. 3220.0*, Australia.

ACCI (2005). Offshoring a viable option for some companies, Australian Chamber of Commerce and Industry Website, Last accessed 10 September 2007, www.acci.asn.au/text_files/issues_papers/Trade/June%2005%20-%20Offshoring.pdf

AHIT (2007). Medicare, NeHTA and your Privacy, *Australian Health Information Technology*, Retrieved September 10, 2007, from http://www.australianit.news.com.au/story/0,24897,22318001-5013040,00.html

ALGA(2005).Planning for an ageing community, *Australian local government population ageing action plan 2004 – 2008*, Retrieved September 5, 2007, from http://www.alga.asn.au/policy/healthAgeing/ageing/plan/

Australian Government (2005).National Strategy for Aging Australia, 2005, *Australian Government, Department of Health and Aging*, Retrieved September 10, 2007, from http://www.health.gov.au/internet/wcms/Publishing.nsf/Content/ageing-ofoa-agepolicy-nsaa-nsaabk4.htm,

AusGov (2007).Doctors and Nurses, Department of Immigration Website, Last Accessed 10 September 2007, http://www.immi.gov.au/skilled/medical-practitioners/index.htm

BAN(2007).Body Area Networks, Fraunhofer, Retrieved September 9, 2007, from http://www.ban.fraunhofer.de/index_e.html

CBS (2004).CBS Study looks at benefits and pitfalls of going offshore, *CBS Media Release*, Retrieved September 9, 2007, from http://announce.curtin.edu.au/release2004/C17704.html.

Churches T (2003). A proposed architecture and method of operation for improving the protection of privacy and confidentiality in disease registers, *BMC Medical Research Methodology*, 3(1), 1-13.

Frost and Sullivan(2005). Australia RFID applications, *Frost and Sullivan Website*, Retrieved September 10, 2007, from http://www.frost.com/prod/servlet/market-insight-top.pag?docid=49019419

Glascock AP, Kutzik DM (2000).Behavioral Telemedicine: A New Approach to the Continuous Nonintrusive Monitoring of Activities of Daily Living, *Telemedicine*, 6(1), 33-44.

Hayes TL, Pavel M, Schallau PK, Adami AM(2003). Unobtrusive Monitoring of Heath Status in an Aging population, 2-5.

HealthConnect (2007a). Introduction, HealthConnect, Retrieved September 10, 2007, from http://www.healthconnect.gov.au/internet/hconnect/publishing.nsf/Content/intro

HealthConnect (2007b). What is happening, HealthConnect, Retrieved September 10, 2007, from http://www.healthconnect.gov.au/internet/hconnect/publishing.nsf/Content/whats-happening-1lp

Hovenga EJS, Garde S, Carr T, Hullin CM (2007). Innovative approaches and processes for capturing expert aged care knowledge for multiple purposes, *Electronic Journal of Health Informatics*, 2(1):e5, 1-13.

Koch T, Jenkin P, Kralik D (2004). Chronic illness Self Management: Locating the self, *Journal of Advanced Nursing*, 48, pp 484 -492, EBN Online, BMJ Publication.

Moor, J.H. (1997). Towards a theory of privacy for the information age, *Computers and Society*, 27(3), 27-32.

Mulligan E, Rogers WA, Braunack-Mayer A (2007).Research within the privacy regulations: Problems and solutions for database custodians, *Electronic Journal of Health Informatics*, 2(1):e7, 1-5.

NEHTA (2007). National e-health transition authority, Nehta Website, Retrieved September 10, 2007, from http://www.nehta.gov.au/

Ohta S, Nakamoto H, Shinagawa Y, Tanikawa T(2002). A Health Monitoring System for Elderly People Living Alone, *Journal of Telemedicine and Telecare*, 8(3), 151-156.

OpenEHR (2007). Research and Development based on openEHR, Retrieved September 20,2007, from http://www.openehr.org/projects/t_projects.htm

Ozanne E, Borowski A, Encel S (1997). *Ageing and social policy in Australia*, Cambridge University Press.

Philipson G, Roberts J (2007). Caring for the future: the impact of technology on aged and assisted living, *Electronic Journal of Health Informatics*, 2(1): e1, 1-9.

Productivity Commission (2005). *Economic implications of an ageing Australi*a. Productivity Commission, Canberra.

Ride (2007). Ride D.2.1.1 – Current European practices in providing interoperability in e-health domain: HEALTHCONNECT and NEHTA (Australia), *Information Society Technologies*, Retrieved September 10, 2007, from http://www.srdc.metu.edu.tr/webpage/projects/ride/deliverables/RIDED.2.1.1%20-%20CurrentPracticesAustralia.doc

Rowland DT (1991). *Ageing in Australia*, Melbourne, Longman Cheshire.

Soar J, Yuginovich T, Whittaker F (2007).Reducing avoidable hospital admissions of the frail elderly using intelligent referrals, *Electronic Journal of Health Informatics*, 2(1):e3, 1-13.

Standards (2007). HL7 standard implementation guidelines, Retrieved September 9, 2007, from http://e-health.standards.org.au

Tavani, HT (2004). *Ethics and technology: Ethical issues in an age of information and communication technology*, Wiley: Hoboken, NJ

Wiebell H (2005). RFID Should Be Restricted by the State, *Information Age*, 14th December, Retrieved September 9, 2007, from http://www.infoage.idg.com.au/index.php/id;1222880574;fp;4;fpid;675408222

PATIENT-CENTRED LABORATORY VALIDATION USING SOFTWARE AGENTS

John McGrory, Frank Clarke
Dublin Institute of Technology, Dublin 8, Ireland
john.mcgrory@dit.ie, frank.clarke@dit.ie

Jane Grimson
School of Computing, Trinity College Dublin, Dublin 2, Ireland
jane.grimson@tcd.ie

Peter Gaffney
Adelaide & Meath Hospital, Incorporating the National Children's Hospital (AMNCH), Tallaght, Dublin 24, Ireland
peter.gaffney@amnch.ie

Keywords: Clinical laboratory validation, patient-centred, patient-focused, agents, computer-interpretable-guideline (CIG), guidelines and protocols.

Abstract: Guidelines are self-contained documents which healthcare professionals reference to obtain knowledge about a specific condition or process. They interface with these documents and apply known facts about specific patients to gain useful supportive information to aid in developing a diagnosis or manage a condition. To automate this process a series of Standard Operating Procedures (SOP) and workflow processes are constructed using the contents of these documents in order to manage the validation flow of a patient sample. These processes decompose the guidelines into workflow plans, which are then called using condition triggers controlled by a centralised management engine. The software BDI agent offers an alternative dynamic which more closely matches the modus operandi of narrative based medical guidelines. An agent's beliefs capture information attributes, plans capture the deliberative and action attributes, and desire captures the motivational attributes of the guideline in a self-contained autonomous software module. Agents acting on behalf of guidelines which overlap and interweave in similar domains can collaborate and coordinate in a loosely coupled fashion without the need for an all encompassing centralised plan.

1 INTRODUCTION

The primary role of a clinical laboratory is to support frontline healthcare professionals who are licensed to deal with patients (McLoughlin, 2006). Their function is to accept appropriate patient samples, analyse them, and report their findings back to the ordering clinician (Marshall et al., 1995). The reported results are not considered a diagnosis, but used by clinicians to deliver patient specific care. In the majority of cases these results are use to aid in planning treatments, quantifying medication amounts and monitoring patient responses, which could all have a detrimental affect on the patient if incorrect information was used (Witte et al., 1997). Therefore, the single most important activity performed by the laboratory technologists is to ensure their generated results are valid and plausible for the specific patient from whom the sample was taken.

Clinical guidelines are condition focused documents through which domain specific aims, goals, procedures, plans and normal reference ranges are disseminated to healthcare professionals. The purpose of these documents are to guide the reader, and streamline activities around a particular medical condition or process using evidence based supportive information. When a clinical or laboratory guideline is developed by an expert group they focus on best practice for the specific condition or process. They include all relevant knowledge, logic and motivational aspects they deem necessary to adequately describe the domain.

Clinicians and laboratory technologists care for patients not diseases or processes, therefore it is their responsibility to filter through these guidelines acting on a patient's behalf. They must try interface with these documents, to make use of the maximum decision-making support for healthcare delivery based on the known facts about their individual patient. To automate this process of searching through guidelines on a patient's behalf, the laboratory technologists in association with the clinicians construct a series of Standard Operating Procedures (SOP), and workflow processes using the contents of these documents in order to manage the validation flow of a patient sample. This is accomplished by decomposing each guideline in to a series of separate workflow activity paths. Then develop a set of centralised management rules to link these activities based on the presented patient data. However, these procedures are not truly patient-centred but process-centred. The guidelines knowledge, logic and motivation can no longer be accessed as a standalone resource, but as a series of workflow triggers managed by a centralised software package, which no longer resembles the author's guideline. This process is fundamentally different to the true operation of a medical guideline, where guidelines are used to provide supporting information based on their holistic view of the domain, rather than a series of linked activities relating to a process. So is there an alternative approach where the process can be distributed and the guidelines retain logic, knowledge and motivation as a standalone self-contained unit.

Agent oriented architectures operate on similar principles to elements found in human decision-making by combining attributes (beliefs), methods (plans) and desires (goals). The BDI agent approach in particular is based on the principle of a belief capturing the informational attributes, the desire capturing motivational attributes and the intention capturing the deliberative attributes of an agent (Rao et al., 1995). Therefore, agents can be considered self-contained knowledge sources (KS), with a social communication interface and have the ability to act autonomously, or as part of a larger group. In research completed by the authors it was shown that a software agent can successfully capture and be encoded with the knowledge, logic and motivation of a guideline (McGrory_a et al., 2008). In additional research completed by the authors it was shown that although agent communications provide a facility to transmit data between agents, it is also used to provide a social and collaborative aspect (McGrory_b et al., 2008). This allows the separate agents work in groups and collaborate on shared goals. This later research also demonstrated that agent communication was capable of being adapted to comply with a medical standard for communication (i.e. CEN ENV 13606-4:1999).

The thrust of this paper is to illustrate that software agents offer an alternative approach to reproduce the function of medical guidelines than the more commonly used centralised approaches. This paper also presents an operation of a framework which allows these agents coordinate and collaborate to validate a patient sample in a distributed fashion, without the need for a centralised all encompassing plan.

2 ANALYSIS OF GUIDELINE REPRESENTATION FORMATS

The traditional approach to combining separate bodies of knowledge (such as guidelines) together is to decompose the knowledge and logic into separate workflow activities and link these activities together using a centralised inference engine. Three commonly used techniques are rule base, direct coupling or blackboard systems.

The rule-base approaches are designed around a nodal tree, where expert knowledge in the form of a workflow activity is the branch, and the selecting of a particular branch at each node is based on patient information or process data. Although selecting rules based on presented facts during execution can be indicative of an illness, the rules which link them directly to a diagnosis do not reflect anything deeper than a casual understanding of human physiology. These systems are centralised and the original guideline knowledge is now absorbed within a labyrinth of rules.

Direct coupling architectures are made up of a group of separate expert knowledge modules. Each expert knowledge module contains local storage, a KS and a control switch to link to the other software modules according to their data-flow requirements using a direct call or link (Corkill, 2003). Complications arise when specific modules are subject to change and/or when the ordering of module control switching cannot be determined until run-time (Kavanagh et al., 2002). As the system expands and evolves the links change and the process becomes unwieldy and unmanageable. In addition to the aforementioned issues, the direct coupling model does not provide a clear

representation of the overall problem, and there is nothing more than relationship links used.

The blackboard model is based around three components: KS, control element and the blackboard (Turban et al., 2005). The KS is an expert at solving specific elements of the overall problem. The blackboard, acts as a central repository for data, partial solutions and control information. The blackboard also acts as a communication medium for the transfer of information, and a KS triggering mechanism. The control element directs the problem-solving process by allowing KS's to respond to blackboard changes, and it selects the most appropriate KS to be executed next, as shown in Figure 1. After completing a task the KS reports back to the blackboard and returns control to the control element. KS's are not aware, and cannot communicate with other KS's directly. They know nothing about the other experts (e.g., what parameters they use, what processes they perform, or what services they provide). The blackboard architecture tends to be a labyrinth of different configurations, levels of abstraction, and partial solutions which are orchestrated to provide a flexible problem solving mechanism. The blackboard system eliminates the communication issues raised by the directly coupled monolithic model, and gives a representation of the problem to be solved to all participants. But the blackboard does not have the capacity to indicate how group members can collaborate to solve a problem, but can only select from partial solutions it already possesses.

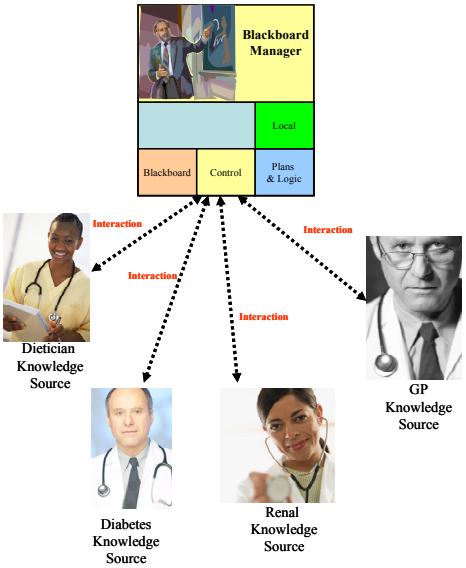


Figure 1: Blackboard Management Communications.

3 DESIGN OF A GUIDELINE AGENT

The agent approach is based on the principle that each agent can represent a single guideline. It captures all the guidelines knowledge, logic and motivation. In addition to this the agent has a rich communication facility where data and social interaction between separate agents can take place. But how does each agent know what the other is doing since there is no centralised all encompassing plan? If the blackboard is simply a repository of information, although layered to some degree, there is no absolute necessity for it to be in a single location. Therefore, it is possible to replicate a copy of the blackboard within each Autonomous Socialising Knowledge agent (ASK-agent) as shown in Figure 2.

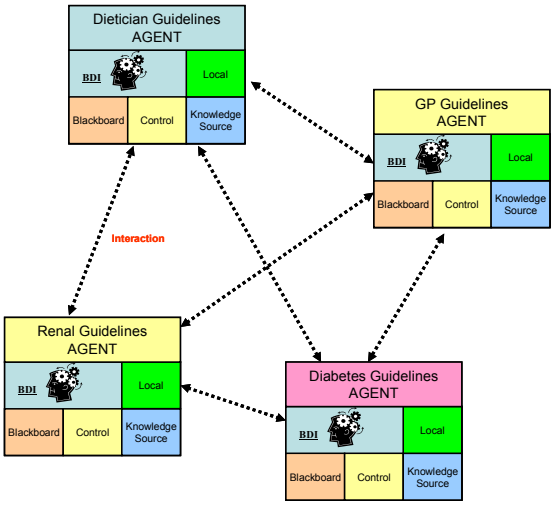


Figure 2: Autonomous Socialising Knowledge agent model.

Each ASK-agent now contains a localised blackboard, knowledge source, localised beliefs, localised control and its own inference engine. The retention of the motivational component of the guideline within the ASK-agent is fundamentally different to the centralised approach of other systems. This allows the ASK-agent to act autonomously on behalf of the guideline in a self-contained capacity. When patient specific information is presented to the individual agents, they have the ability to apply their encoded knowledge and logic, and provide a supportive response based solely on that information. Using this approach an ASK-agent module can make use of the maximum supportive response from the other

separate ASK-agent's based on the known facts about the individual patient. By providing a framework which allows separate ASK-agent broadcast supportive communications to each other, the agent approach offers the opportunity for the data to be validated in a patient-centred fashion. But how can these separate, autonomous, self-contained ASK-agent modules share data, work in groups or collaborate to solve a problem.

3.1 Agent to Agent based Activity

A theory which can aid issues relating to collaborating guidelines is Activity Theory (AT) which emanated through the social sciences. AT focuses on the collaborative nature of separate autonomous systems such as individuals (Engestrom et al., 1999), on which agents are based, and have the capability to perform certain tasks as part of a group. Agents synthesise human decision-making through their goal, plan and belief elements, but do not explicitly detail how they can socialise or collaborate. AT in itself does not provide an output which can be exactly transposed into computer software, but does provide a useful framework based around interfacing interaction and collaboration of software modules. These interfaces can be used to develop an increased sense of interaction and collaboration ability in autonomous modules using a software program independent approach.

The structure of human activity according to Engestrom can be compartmentalised using *rules*, *community*, *subject*, *object*, *division of labour* and *instruments*. AT is an iterative process where an activity is developed from a simple low level activity to a higher level activity.

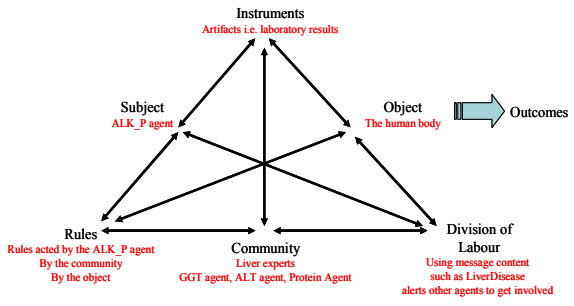


Figure 3: Low level activity.

In terms of medical guidelines the iterative process dynamic exists by virtue of the design of guidelines and their focus on a condition, disease or organ. Consider for example, a low level implementation activity being the guideline behind the validation of a single analyte result, say Alkaline Phosphates as shown in Figure 3. A higher level implementation is where the result is combined with some other single analyte results, such as Bilirubin and GGT, to perform and aid in the reporting of a Liver Function Test as shown in Figure 4. The Liver Function Test is then part of a higher level suite of tests for other medical disorder classifications.

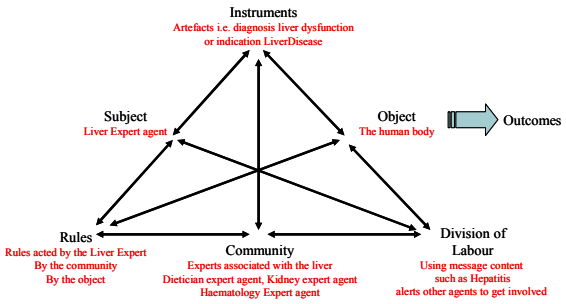


Figure 4: High level activity.

One guideline does not cover the whole body, but more specifically focuses on an abstract conceptualisation of body components (e.g., liver function in a group of male diabetes patients). Another guideline relates to the same body component but from a different abstract conceptualisation viewpoint (e.g., kidney function in a group of male diabetes patients). Although the two guidelines are separate autonomous documents they are linked by virtue of their domain of discourse. Therefore, a link between different guidelines already exists within the guideline document itself. The overlapping knowledge is provided in two main forms. The first is in the form of similar domain knowledge that uses alternative inference mechanisms in order to derive a result (i.e. both statistical and rule-based inference engines being able to validate the same result). The second is in the form of overlapping knowledge which observes different viewpoints of the same domain. For example, the kidney filters toxins from the blood passing it to the urinary tract. As the kidney is such an integrated organ in the body there are many guidelines describing its operation from different viewpoints such as blood filtering, urinary tract, autoimmune disorders etc. Using this approach the organ disease or condition is described from different viewpoints through various guidelines. Each guideline describes different knowledge, logic and motivational aspects associated with the organ. Therefore supportive information can be exchanged between these guidelines in order to aid in

describing the operation of the organ, or in the validation of a sample result.

3.2 ASK-agent Model

To utilise this overlapping knowledge link a social structure was developed to manage the interfacing between agents. This social interface took the form of a mandatory set of searchable service descriptions, beliefs and actions. The service descriptions (i.e. *Name, Type, Ownership, GuidelineReference, InformationNeeded, ValidationType, EndResultType, Ontology* and *Language*) permitted each agent to be located within the agency platform through the Directory Facilitator (DF) (a feature of the Foundation for Intelligent Physical Agents (FIPA) standard offering searchable *goldenpages* facility to locate agents) (McGrory_a et al., 2008). The beliefs (i.e. *CurrentlyValidating, PlausibilityScore* and localised blackboard) permit the ASK-agent to interact with other group members. The actions relate to automated responses the ASK-agent must return to other agents when queried (e.g. *CurrentlyValidating*), and the sending of information to other agents it believes should be reported (e.g. it determined the presence of liver disease during its deliberation). Therefore, each ASK-agent only needs to know its overlapping neighbours, which it can find and interact with using the agent platforms DF and message passing. With access to supportive and overlapping knowledge it is not necessary to have a single all-encompassing rule set to manage the ASK-agents interaction.

The fundamental concept of the ASK-agent system proposed in this paper is to allow components to collaborate and share supportive information without having to explicitly disclose their position as part of the large encompassing community. An ASK-agent does not need to identify exactly what every other agent is doing; only what its neighbours (i.e. neighbours it interfaces with) are doing. To illustrate this point further and demonstrate some boundaries, consider the example of a jigsaw with 500 pieces. A jigsaw piece has two discrete dimensions: the irregular shaped edge containing four sides, and the image printed on the face. To solve the puzzle, a person directly matches individual jigsaw pieces onto the jigsaw image, say the image shown on the box. Each piece is identified using the image on its face and placed in the appropriate position. This method requires a view of the whole system to be presented before starting, but involves no greater skill than straightforward pattern matching. An alternative approach is to use a combination of the localised image on the face of the piece and its four corners to match it to a suitable neighbour (i.e. matching the shape of the pieces together). Jigsaw assembly using these interfaces do not require the full picture to be known. Using the jigsaw example as a solution metaphor, the heart and lungs image depicted in Figure 5 is a symbolic representation of the heart and lungs as a whole, not just the image it represents. The agent was not intended to be a large all encompassing structure, but a group of loosely coupled autonomous expert knowledge sources (represented by each jigsaw piece) which could be readily and easily interfaced with as shown in Figure 5. The ASK-agent only needs to know its neighbours (i.e. the expert it interfaces with), in a similar way the jigsaw piece only needs to know another piece with similar edge profile and compatible image, not the whole picture. The ASK-agent does not need to know anything about any other piece of the jigsaw only its interfacing neighbours. The interface can be considered the ontology, overlapping facts, common laboratory results and various viewpoints of the universe of discourse. This is analogous to the jigsaw edge shape profile.

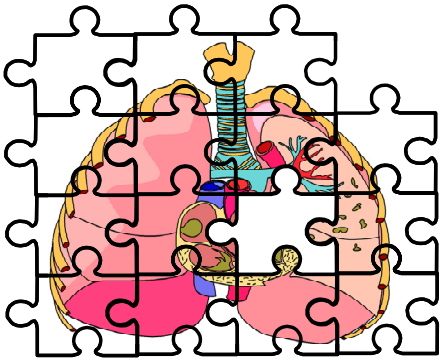


Figure 5: Jigsaw metaphor representing agent components of the heart and lungs.

4 CONCLUSIONS

This research demonstrates the agency approach offers a facility to manage and interface with medical guidelines electronically, in a similar modus operandi to original guideline documents. This is because of the synergy between the knowledge base, plans, decisions, action, goals and the self-contained nature components between guidelines and agents.

Table 1: Summary of centralised and ASK-agent approach.

Element	Centralised Approach	ASK-Agent
Multi-ontologies	No	Yes
Processing	Centralised	Distributed
Requirement for overlapping knowledge	None	Required to provide the links.
Addition, altering or removal of guidelines from the system	Any changes require centralised inference to be recompiled.	Each ASK-agent is independent and loaded separately.
System resilience	None	Yes, all ASK-agents have a copy of the blackboard.
Independent accessible knowledge	None. All access to information through the centralised engine.	Yes, all ASK-agents are independent. Information accessed via message passing.
Clinician having access to specific guideline knowledge	A clinician cannot access knowledge directly.	A clinician can access each ASK-agent via a message and directly access the specific guideline knowledge.
Method of collaboration	Direct links using the centralised engine.	Using the mandatory beliefs, action and descriptions in the agent platforms Directory Facilitator.

The agents can be encoded to reproduce the beliefs, desires and intentions of the narrative guideline and act accurately, faithfully and autonomously on behalf of that document. This body of knowledge and logic can then be interfaced with, whenever that information needs to be accessed. The addition of activity theory and in particular the iteration model concept showed that the guideline documents already contain aspects that link them together. Using these links and the developed social communication the ASK-agents can locate, access, communicate, collaborate and coordinate activities between each other. This allows supportive information exchanges to be completed between separate expert agents about an individual patient, without the need for an all encompassing centralised plan. In cases where there is an inconsistency in held patient specific information, this agent approach offers an advanced, robust and efficient patient centred validation alternative to existing approaches. However, if overlapping knowledge between guidelines is not available the links created using this approach are not present and the separate guidelines are standalone islands of information. The guidelines knowledge, logic and motivations are still accessible as a standalone entity, but other agents would need to be created to provide the links. Developing a system using the latter approach still permits distributed processing to be accomplished, but not without a source of knowledge to provide the links. A summary of the differences between the centralised and ASK-agent approach are given in Table 1.

REFERENCES

Corkill D., 2003. *Collaborating Software Blackboard and Multi-Agent Systems & the Future*. Accessed via www.cs.umass.edu, Proceedings of the International Lisp Conference published in October 2003.

Engestrom Y., Miettinen R., Punamaki R. *Perspectives on Activity Theory*. Published by Cambridge University Press January 1999. ISBN13: 9780521437301.

Kavanagh M., Price S., "*The quest for a computerized guideline standard: The process, its history, and an evaluation of the most common and promising methods used today*", Capstone Project documentation 2002.

Marshall W.J., Bangert S.K., 1995. *Clinical Biochemistry, Metabolic and Clinical Aspects*, ISBN 0 443 043 418, published in 1995 by Churchill Livingstone.

McGrory_a J., Grimson J., Clarke F., Gaffney P., "*Agents representing medical guidelines*", HEALTHINF 2008 International Conference on Health informatics, Funchal, Madeira, Portugal, January 2008.

McGrory_b J., Grimson J., Clarke F., Gaffney P., "*Communication of medical information using agents*", HEALTHINF 2008 International Conference on Health informatics, Funchal, Madeira, Portugal, January 2008.

McLoughlin V., Millar J., Mattke S., Franca M., Jonsson P., Somekh D. and Bates D., 2006. *Selecting indicators for patient safety at the health system level in OECD countries.* International Journal for Quality in Health Care 2006 18(Supplement 1):14-20.

Rao A., Georgeff M., 1995. *BDI Agents: From Theory to Practice,* proceedings of the first international conference on multi agent systems (ICMMAS-95) San Francisco USA 1995.

Turban E., Aronson J.E., Liang TP., *Decision Support Systems and Intelligent Systems.* published in 2005 by Pearson Prentice Hall, Seven Edition, ISBN 0-13-123013-1, Chapter 10, pgs 538-570.

Witte D.L., VanNess S., Angstadt D., and Pennell B.J., 1997. *Errors, mistakes, blunders, outliers, or unacceptable results: how many?* Published by Clinical Chemistry 1997; Vol:43 pgs 1352-1356.

ON DESIGNING AN EHCR REPOSITORY

Petr Aubrecht, Kamil Matoušek and Lenka Lhotská
The Gerstner Lab, Czech Technical University in Prague, Technická 2, 166 27 Praha 6, Czech Republic
matousek@labe.felk.cvut.cz, aubrech@labe.felk.cvut.cz, lhotska@fel.cvut.cz

Keywords: EHCR, databases, multiagent systems.

Abstract: In an ongoing project, a pilot study and implementation of repository design for electronic home care records (EHCR) is described. Electronic home care record is based on the idea of electronic health record, however it also satisfies additional information and functionality requirements specific for home care. The design is based on the home care data and service model (K4Care model). First we analyzed the problem and decided about the platform, storage technology, cooperation with other parts of the system being developed, and basic structure of the EHCR. Then we focused on the design of data storage and transformation of the K4Care model into a database structure. Finally cooperation between the database and multiagent system is proposed.

1 INTRODUCTION

An electronic health record (EHR) is a distributed personal health record in digital format. The EHR should provide secure, real-time, patient-centric information to aid clinical decision-making by providing access to a patient's health information at the point of care. An EHR is typically accessed on a computer or over a network. An EHR almost always includes information relating to the current and historical health, medical conditions and medical tests of its subject, thus representing a longitudinal collection of information for and about patients. In addition, EHR should contain data about medical referrals, medical treatments, medications and their application, demographic information and other non-clinical administrative information.

Although there are few standards for modern day electronic records systems as a whole, there are many standards relating to specific aspects of EHRs. These include:

ASTM Continuity of Care Record (CCR) is a patient health summary standard based upon Extensible Markup Language (XML); the CCR can be created, read and interpreted by various EHR or Electronic Medical Record (EMR) systems, allowing easy interoperability between otherwise disparate entities.

American National Standards Institute (ANSI) X12 (EDI) is a set of transaction protocols used for transmitting virtually any aspect of patient data. It has become popular in the United States for transmitting billing information, because several of the transactions became required by the Health Insurance Portability and Accountability Act (HIPAA) for transmitting data to Medicare.

CEN, the European Committee for Standardization, was founded in 1961 by the national standards bodies in the European Economic Community and EFTA countries. Now CEN is contributing to the objectives of the European Union and European Economic Area with voluntary technical standards which promote free trade, the safety of workers and consumers, interoperability of networks, environmental protection, exploitation of research and development programs, and public procurement. CEN - CONTSYS (EN 13940) is a system of concepts to support continuity of care. CEN - EHRcom (EN 13606) is the European standard for the communication of information from EHR systems. CEN - HISA (EN 12967) is a services standard for inter-system communication in a clinical information environment.

DICOM is a heavily used standard for representing and communicating radiology images and reporting

Health Level Seven (HL7) is one of several ANSI-accredited standards developing organizations operating in the healthcare arena. HL7 messages are used for interchange between hospital and physician record systems and between EMR systems and practice management systems; HL7 Clinical

Document Architecture (CDA) documents are used to communicate documents such as physician notes and other material.

Integrating the Healthcare Enterprise (IHE), while not a standard itself, is a consortial effort to integrate existing standards into a comprehensive best-practice solution

ISO TC 215 has defined the EHR, and also produced a technical specification ISO 18308 describing the requirements for EHR Architectures.

OpenEHR represents a next generation public specifications and implementations for EHR systems and communication, based on a complete separation of software and clinical models.

Various factors involving the timing, the right players, market history, utility, governance play a key role in the overall enrichment of the standard and certification development. The standardization and certification even though seem to bring uniformity in the EMR development; they do not guarantee their acceptability and sustainability in the long run.

The core of any EHR system is a data repository that is usually realized by a database system. Types of databases include the following ones (Bontempo & Saracco, 1995): hierarchical, network, relational, and object oriented. The hierarchical and network databases represent older forms that are not used in newer applications. Relational database departs significantly from those two types and is the most common form of database today. Relational databases are constructed using tables instead of tree and network structures. The tables do not specify how to retrieve the required data or navigate through predefined path. Object oriented database is the most recent approach to database management. The object oriented database structure is derived from object oriented programming and has no single inherent database structure. The structure for any given class or type of object can be anything a programmer finds useful. Furthermore, an object may contain different degrees of complexity, making use of multiple types and multiple structures.

In the next sections we will describe individual issues related to EHCR design in more details, namely persistence and structure of EHCR.

2 SCHEMA OF PERSISTENCE

In programming, persistence refers specifically to the ability to retain data structures between program executions, such as, for example, an image editing program saving complex selections or a word processor saving undo history.

This is achieved in practice by storing the data in non-volatile storage such as a file system or a relational database or an object database. Design patterns solving this problem are container based persistence, component based persistence and the Data Access Object model. When first introduced, the idea was that persistence should be an intrinsic property of the data, in contrast with the traditional approach where data is read and written to disk using imperative verbs in a programming language. This emphasis has largely disappeared, resulting in the use of persist as a transitive verb: On completion, the program persists the data.

2.1 Storage Technology

For persistence, there were two ways in our case:

- Relational database;
- XML storage.

Relational databases have long history and today are able to store data reliably and efficiently. There exist a big number of available databases both commercial and free.

XML databases are relatively new technology with promising future, allowing natural storage of data with complex structure (e.g. annotated texts). On the other hand, there is no emphasis on reliability. Also effectiveness of query execution (using XPath) is usually low.

Because the content of the database is critically important, we emphasize robustness of the storage. Another important point is amount of data to be stored, thus the ability of the technology to manage huge dataset is crucial. Therefore the chosen technology is relational database.

This requirement also affected the further decision about which database to use. We analyzed advantages and disadvantages of several potential databases. The database we typically use for storage of a big amount of data is PostgreSQL. It provides a rich query language with strong emphasize of the Structured Query Language (SQL) standard. It is proven to be reliable and has a great support for transactions. Its BSD licence is the most free and open. From the administrative point of view, PostgreSQL allows use of several languages for stored procedures including Java, which can be used for simplification of some tasks. For this database there exists a convenient graphical interface. The performance of PostgreSQL is more than sufficient.

Another database server is represented by Mysql. It is known to be very fast for read-only databases. Unfortunately, it has significant drawbacks – limited SQL and transactions supported only in one backend with a loss of performance. Practical experience shows many issues e.g. saving invalid dates, accepting syntax errors, or even unpredictable errors. Its license is not so free, the resulting product must either be of GNU General Public License (GPL) or Mysql must be bought.

There are other databases, which could be used. For example, Firebird, which is free, but with many specifics. Microsoft SQL Server faced several security issues in past and requires a specific operating system.

Big players in the field – Oracle and DB/2 are extremely expensive for our purpose. Although Oracle provides a free version, it is limited in database size.

Based on the analysis we decided to use PostgreSQL. In case of remote database access, we are going to establish a Secure Sockets Layer (SSL) connection to the database to protect the transferred data.

2.2 Overall Schema of Persistence

When we started the project, one of the key questions was the format of messages in the system. The formats discussed were the medical standards HL/7 and ENV 13606 standards. We all decided to implement these standards consistently with their purpose – only for data interchange between this system and external institutions. In detail, this communication will be operated by specialised translation software agents (Hyacinth, 1996). We suggest that import will be manual in this project, because of the small amount of medical data to import and because the implementation is too complex.

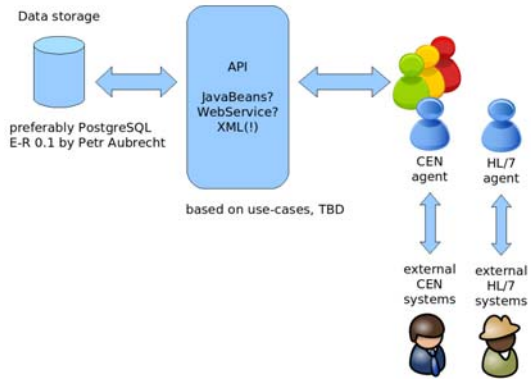


Figure 1: Schema of communication.

The format of messages within the multiagent system – MAS (Wooldridge, 2002) will depend on the concrete content of documents described in the K4Care model (Campana, 2006). The overall schema is in the Figure 1.

3 DATABASE STRUCTURE

3.1 Document Storage

For document storage there were two approaches available:

- Data oriented with all data precisely defined and stored in multiple tables;
- Document oriented with documents seen as an atomic structure with no intervention with its internal structure.

The decision is quite simple – the design of the required documents is an incremental process, documents are analyzed and their detailed structure is prepared. Moreover, we expect in the future more document types, which will lead to modification of the structure. In the first case it requires changes in the database structure and also in the API between agents and data storage.

The decision is then to store particular documents as XML documents without knowledge (from the storage/database point of view) of its internal structure. Together with documents their schema will be stored – at the moment of document storage, the schema must be known.

There is one key requirement on the document schemas – to contain a text part for agents, which will not understand the particular type, to be able to display at least some information to the human client.

3.2 Structure of EHCR

The schema evolved originally from the fundamental report of the project (Campana, 2006), describing the problem domain. We identified data, which are required to be stored and then we suggested a structure holding all the data. Later, a part of the schema, e.g. rights of particular users to some actions/documents, formed the ontology layer. A part of the schema is shared with the ontology layer. Figure 2 shows the schema of the storage.

The objects on the storage side are divided into three groups:

The first group forms EHCR, e.g. description of a patient and all his/her medical documentation. In

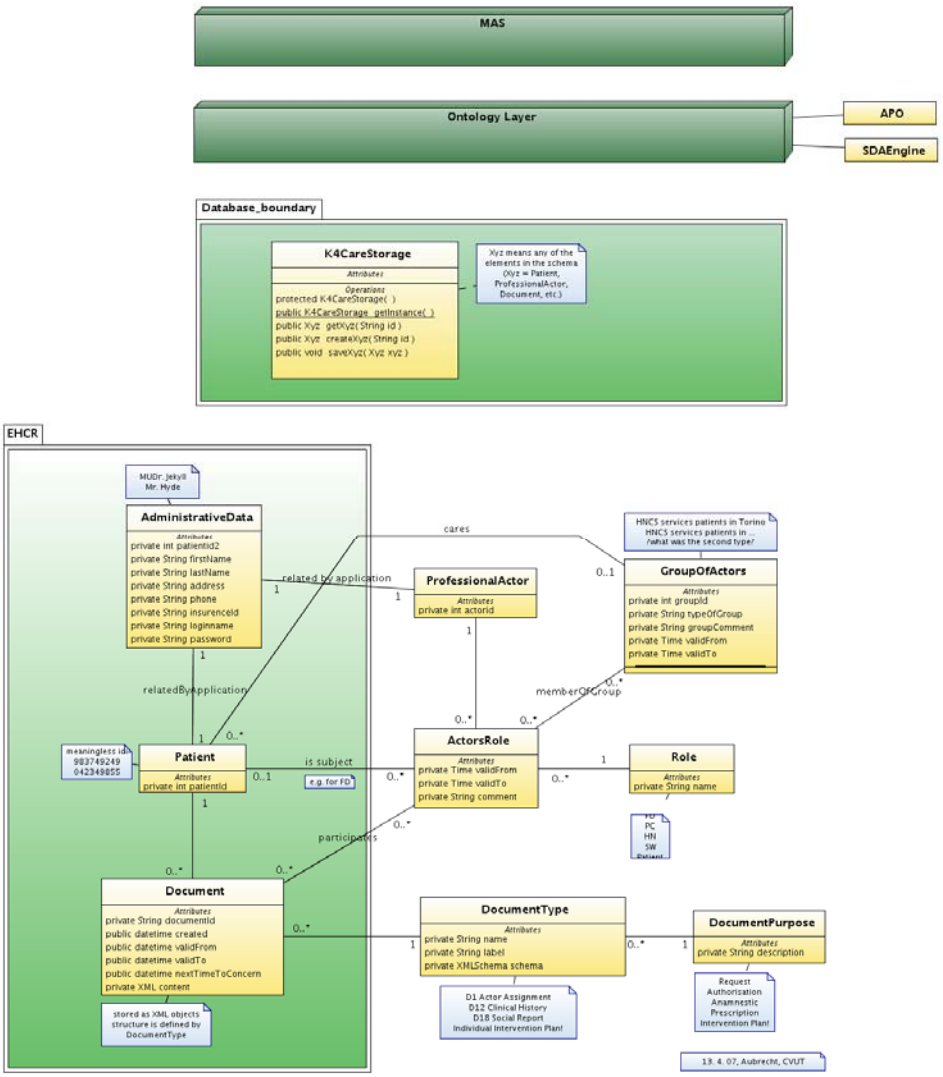


Figure 2: K4Care Model.

the schema this part is grouped into light green area. It consists of administrative data and documents. One of the requirements within the project was to separate these two kinds of information, so the relation between them is not presented in the database, but implemented in Java program by a kind of a cryptic function. These two parts can be also physically separated (in two databases, possibly on two computers) and joined by the program. Documents are stored as XML-structured texts.

The second group of objects, *ProfessionalActor*, *ActorsRole*, and *GroupOfActors*, describes the structures used in the system. Information about professionals is stored in a similar way to the patients – separately from the administrative data. An important part of these objects is dedicated to groups of the professionals and which patients are served by these groups.

The third group of objects, *DocumentType*, *DocumentPurpose*, and *Role*, is a part of schema shared with ontology layer. Information about document types and purposes and about roles of professionals must be stored in both storage (e.g. because it is required to find professionals or documents with a professional in some particular role) and ontology (these information is a part of descriptions of rules performed by the ontology layer).

Documents represent an important part of the whole schema. They are stored in XML format due to the structured content, which can be changed in the future. The corresponding XML schema is stored in the *DocumentType*, so during each save operation, the structure of the XML document can be verified. As a result, the data store does not "understand" the content of the XML documents. Their processing is done by the higher levels of the system. Only a few common attributes are stored explicitly as specific attributes. These attributes are used in document search.

Each document is assigned an attribute *NextTimeToConcern*, which allows launching some actions in the future, like appointments after some period, regular vaccination, etc.

Besides the document data, the EHCR have to store the following, so called *intervention plans*:

- *Formal Intervention Plan* (FIP). It is not going to change for a long time and belongs to a group of diseases, syndromes or symptoms (from ontology layer). FIPs will not be stored in the repository, as they are not connected to any patient. The stored FIPs may be associated to diseases, syndromes and even symptoms, described by codes proposed by medical partners using the International Classification of Diseases version 10. These codes should be used in the formal/individual intervention plan descriptions.
- *Individual Intervention Plan* (IIP) belongs to a single particular treatment of a patient. This kind of intervention plan is valid only for the time of the treatment. It is created on demand and based on a FIP. It will be attached to the EHCR of the patient.

The IIPs will be stored as documents, similarly to any other medical information. The plans including FIPs and IIPs will be expressed using SDA* model (Riaño, 2007). An interpreter and a graphical interface for creating and modifying the plans are under development.

Steps written using the SDA* model will be performed by software agents in the MAS part of the system.

3.3 Multiagent System and Database Cooperation

Cooperation between multiagent system and storage is schematically shown in figure 2. First, between these two parts is an ontology layer, responsible for checking rights and providing semantic services to the MAS.

In the upper part, there is the interface of the storage to the outside components (client side). It can be either ontology layer, as designed in the K4Care project, or directly the multiagent system.

The opposite side uses *K4CareStorage* object as an entry point to the storage and as a controller for creation and modification of all of the available objects. For each object it provides all of the four operations: creation, retrieve, update, and remove.

Objects returned are connected together using pointers to the related object, so the connection using primary and foreign keys (used in the database) is hidden to the client side and thus convenient.

Both the developers of the agents and the developers of the database may need to modify such interfaces. These modified interfaces will be handled carefully.

4 DOCUMENT SCHEMAS

Currently, XML schemas are being created to make clearer the structure of the EHCR documents planned to be used in K4Care. All the documents defined in (Campana, 2006) can be incorporated in EHCR as long as they belong to the patient.

The structure of the most relevant documents is being discussed with the medical partners. Real documents are examined and described in the form of XML Schema. Together with documents are stored intervention plans as a specific type of document (in most cases with *NextTimeToConcern* attribute set, when a next step of the plan should be performed).

In the tables 1 and 2, there are shown for illustration the most relevant service specific and common documents to be implemented in the first stage:

Table 1: Examples of the service specific documents for the first stage.

Document name	Abbreviation and type
MDE scales	D11 – anamnestic multi-dimensional evaluation
Set of forms filled in by the evaluation unit (EU) during the first problem assessment and/or in occasion of the periodical or end-treatment re-evaluation.	
Clinical history	D12 – anamnestic clinical assessment
All the available pertinent clinical information of the patient (HCP) – previous test results, discharge sheets, consultations, previous treatment. It is written by the family doctor (FD) and the physician in charge of the home care (PC); it is read by EU and the other professionals in charge of the patient (according to their competencies in the process of care of the individual patient), by the patient him/herself.	
Physical examination report	D13 – anamnestic physical examination
The report contains signs and symptoms of diseases and/or conditions written by FD, PC, specialist physician (SP); can be updated in any occasion of evaluation. It is read by FD, PC, head nurse (HN); SP and nurse (Nu) in charge of the patient.	
Medical follow-up form	D19 – anamnestic follow-up
It is written by FD or PC during the follow-up activities.	
Nursing follow-up form	D20 – anamnestic follow-up
It is written by Nu or HN during the follow-up activities.	

Table 2: Examples of the common documents for the first stage.

Document name	Abbreviation and type
Actor assignment	D01 – request
Is the information that links an individual action to individual HCP for an action (or series of actions) to be performed.	
Actor confirmation	D02 – authorization
Is the information that declares that the actor knows the assignment and accepts it.	

5 CONCLUSIONS

The initial stage of the design and development of EHCR data repository in the frame of the K4CARE project has been described. Data storage was chosen with respect to the nature of the data – absolute requirement of secure and safe way of handling. For this purpose there were used industrial standards for storage: robust relational database engine with transactions and SQL query language. In case the application will access storage remotely, it can provide SSL connection to the server.

The schema of data stored in the database reflects the needs of medical specialists. Documents themselves are stored in XML format in order to allow further evolution of their structures. Other data is stored in relational form to allow fast search and concurrent access.

The overall schema is general enough to allow cooperation with external partners using different formats and efficient inner communication within parts of the system.

ACKNOWLEDGEMENTS

The research has been supported by the EU FP6 project No. 026968 K4CARE “Knowledge-Based HomeCare eServices for an Ageing Europe” and partially by the GA CR project No. 201/05/0325 “New Methods and Tools for Knowledge Discovery in Databases”.

REFERENCES

Bontempo, C. J., & Saracco, C. H. (1995). *Database Management Principles and Products*, Upper Saddle River, N.J.: Prentice-Hall PTR

Campana, F., et al. (2006). *D01 – The K4CARE Model*, interim report.

Hyacinth, S. N. (1996). Software Agents: An Overview, In *Knowledge Engineering Review*, 11(3):1–40, September 1996, Cambridge University Press.

Riaño, D. (2007). The SDA* Model: A Set Theory Approach. Machine Learning and Management of Health-Care Procedural Knowledge. In *CBMS 2007, 20th IEEE International Symposium on Computer-Based Medical Systems.* IEEE Computer Science Press.

Wooldridge, M. (2002). *An Introduction to MultiAgent Systems*, John Wiley & Sons Ltd.

ORGANIC COMPUTING FOR HEALTH CARE SYSTEMS
Possible Benefits and Challenges

Florian Nafz, Frank Ortmeier, Hella Seebach and Wolfgang Reif
Lehrstuhl für Softwaretechnik und Programmiersprachen, Unversität Augsburg
Universitätsstr. 14, 86135 Universität Augsburg, Germany
ortmeier@informatik.uni-augsburg.de

Keywords: Organic Computing, multi-agent systems, health-care applications, security.

Abstract: Todays health care institutions will undergo major changes in the next two decades. The reason for this is the change of ageing structure in many industrialized countries. In Germany statistics indicate that the costs for health care systems will at least double per person while the number of contributing, working citizens will significantly lower. At the same time average life expectation will rise above 80 years. To cope up with this development adaptations to organization and process of health care are necessary. Typically tasks in stationary health care can be divided in two groups: task which incorporate direct interaction with the patient (care tasks) and tasks which focus on logistics and organization (background tasks). In health care it is not desirable and feasible to reduce efforts in care tasks. So costs and efforts must be reduced within the second group of tasks. This is possible if new paradigms – both in organization and underlying software architecture – are applied. One such paradigm is organic computing. Organic computing aims at systems, which are self-organizing, self-adapting to new challenges and self-optimize during runtime. Such systems can take away a lot of organizatorial work form the staff and thus allow for more and better care without rising budgets. The paper outlines the idea of organic computing as well as opportunities and challenges for applying it in the health care context.

1 INTRODUCTION

It is a widely accepted fact, that cost pressure will make changes in todays health care institutions necessary. Reasons for this development are the changing ageing structure and the rising life expectation of many industrial countries' population. In 2005 there were in average 31.7 persons older than 65 years for every 100 persons in working age. This percentage will rise to 54.2 persons in 2030 and to 70.9 persons in the year 2050 (Statistisches Bundesamt, 2006). At the same time the average per year cost for health care for persons older than 65 years is four times the average cost for younger people (Statistisches Bundesamt, 2007). These two facts together show that costs will dramatically rise while the income will stay almost constant.

It is not feasible to simply cut health care benefits to reduce costs. Therefore it is necessary to provide health care more cost efficient. Humans play – and will always play – the most important role in caring for others. So it is not possible and wanted to "remove" human staff from the caring process. To keep the health care systems affordable and survivable it will be of uttermost importance to support the staff in all tasks, which are not directly related to the patients as much as possible.

This can only be done by providing intelligent infrastructure and background systems, which support the medical staff autonomously whenever possible. A paradigm for construction of such systems is Organic Computing (VDE/ITG/GI, 2003). It allows for constructing flexible systems, which dynamically (re-)configure and autonomously self-adapt to changing tasks and requirements.

In this paper a brief overview on Organic Computing is given in Sect. 2. Sect. 3 shows chances and possibilities of applying Organic Computing in health care on a little example. Open questions for research and challenges are discussed in Sect. 4. Sect. 5 concludes this position paper.

2 ORGANIC COMPUTING TECHNIQUES

The complexity of modern systems has grown immensely in the last years. In particular the trend to replace hardware solutions with software components has made the software become (a) more and more im-

portant for the functionality and (b) more and more complex to develop. A good example is a modern (digi-)cam. Five years ago most cameras where using analogous media to record pictures. Extra functionalities like (visual) noise suppression or image stabilization where implemented by using specific media or by adding extra hardware components. Today such functions are usually implemented in software and can relatively easy be transfered from one camera to another. This led to very complex software and this process is now reaching a saturation level. To further enhance the capabilities of both – the system in general and the controlling software in particular – new approaches are necessary to deal with the growing complexity. One such approach is Organic Computing (Müller-Schloer, 2004).

The Organic Computing Paradigm. Organic Computing (Müller-Schloer et al., 2004; VDE/ITG/GI, 2003) is an extension to Ubiquitous and Autonomic Computing (Kephart and Chess, 2003). The core goal of Organic Computing is, that future (computing) systems should be able to dynamically adapt to changes in requirements, to automatically detect and neutralize component failures and to continuously optimize themselves for better performance. Such capabilities are called self-adaptation, self-healing and self-optimization. The hope is, that systems with self-x capabilities are by far superior to standard systems in terms of possible functionalities, availability, reliability and effort for maintenance. At the same time self-organization mechanisms allow for easier construction. As not every single scenario has to be anticipated at design time. New scenarios and requirements will be detected by the system during runtime and self-adaptation mechanisms will reconfigure the system to this new challenges. So the systems behave from an external point of view very much like living beings. They monitor their environment, reflect upon changes and adapt to new situations. Therefore such systems are called “Organic Computing” systems.

It is often useful and possible to cope mechanisms and techniques from biology/sociology for the design of self-X algorithms and systems – so called bio-inspired algorithms. One example is a (self-)protection algorithm which function like the human immune system (Pietzowski et al., 2006). But there also exists a variety of other algorithms, which are based on more traditional approaches.

Technically, implementations of Organic Computing systems often contain one part which is responsible for delivering the wanted/intended functionality and one part which constantly monitors the environment and – if necessary – controls/changes the functional part. There already exists a broad variety of methods (Seebach et al., 2007; Richter et al., 2006), middlewares (Trumler et al., 2004; Trumler, 2006) and analysis tools (Güdemann et al., 2007; Güdemann et al., 2006) for design, analysis and construction of such systems.

Application Domains. Organic Computing is typically applied to software-intensive embedded systems. Example domains are large networks of sensors, traffic control systems or production automation. One example could be a vision of the next generation of production cells. Assume a production cell which should process workpieces following a given specification. The functional part of this system may be a set of robots which perform certain tasks – e.g. drilling holes, inserting screws and tightening screws – on workpieces and a set of transport units, which transport workpieces from station to station. If new types of workpiece are to be processed, then the robots must be reprogrammed (or these new workpieces must have already been anticipated during system design). If a single robot/component of the system fails, then production will often come to a stand still.

If in contrast this system is designed in an organic way, then each robot will be enhanced with a supervising (software) component. This software not only monitors the supervised robot, but also decides/plans what tasks the robot will perform an from where workpieces are taken resp. where workpieces are to be placed next. If new workpieces – with a new specification about how they are to be processed – are given to the production cell, then the supervisors try together to find a reconfiguration to achieve the new goal.

Assume for example a cell that consists of three robots, which are all capable of drilling holes, inserting screws and tightening them and a set of autonomous carts for transportation, then an (initial) configuration could state that one robot only drills, one only inserts screw and one only tightens the screws. The workpieces will be transported in badges from robot 1 to robot 2 to robot 3. The supervisors constantly monitor the functionality of the robots and the type of workpieces which are to be processed. If now for example the drill of robot 1 breaks, then the supervisors will search for a new configuration to subsume production. One solution is to let robot 2 do the drilling, assign robot 1 to inserting screws and change the transportation routes accordingly. This is called self-healing. If a a new type of workpiece is to be processed; e.g. workpieces now only need holes (and no

screws at all). Then the supervisor will recognizes this and reconfigure all robots to drill (self-adaptation)[1].

The technical basis for such effects is (i) the capability of the supervisors to reflect their controlled subsystems capabilities, (ii) to interact/communicate with each other and to (iii) use a common language about goals and capabilities. In the last years there have been big advances in building Organic Computing systems. There exists now a number of working middlewares, organic architectures and intelligent reconfiguration algorithms. They have also been successfully applied to various technical scenarios. But there exist hardly any application to systems in which privacy and individual user-trust play an important role.

3 ORGANIC COMPUTING IN HEALTH CARE

How can Organic Computing help for health care? What are the challenges, which have to be met to allow for using Organic Computing in health care? These question can be best answered, if an example scenario is taken into account.

A possible example scenario may be support and comfort functionality in future rehabilitation centers. This scenario is only one selected example of a broad class of background support systems both inside and outside the domain of health care.

An Example. The scenario assumes, that during the next decade patients as well as medical staff and doctors will be equipped with mobile devices, which can form ad-hoc-networks. The topology of these networks will be continuously changing as the device are carried around by their possessors. Stationary devices like ergo-meters or central services will also be equipped with wireless communication and can also participate in the network.

Typically patients in a rehabilitation clinic arrive with a fixed treatment concept. This concept states a number of treatments (e.g. physio, massage, etc.) which have to be applied to the patient in certain intervals (e.g. twice a week). Periodically – often every two weeks – these concepts may be changed or updated according to the physical state of the patient. Using the treatment concept and the capacities of the staff and installations as input data, a central administration service creates a treatment plan for the patients. This plan schedules treatments to certain time slots and staff/installations. Similarly work plans for the staff are created. This process is relatively complex and requires intensive planning. It is usually done on a weekly basis. The plans are typically printed and handed out to patients and staff.

This process is performing adequately while there are no unpredicted events. In reality there typically occur a lot of disturbances – like patients coming to late to their treatments, illness of staff or malfunction of some equipment. The consequence is that the plans must be adapted. This is often not possible or only if the disturbance can be foreseen (like illness of staff). As a consequence in most cases treatments will either be canceled or at least postponed for several days. The root of this problem lies in the static nature of this process and the central planning.

Benefits of Organic Computing in Health Care. Better results can be achieved if an Organic Computing systems is implemented for this task. In an organic approach, the treatment plan will not be calculated a priori by a central administration. It will rather evolve by interaction of all agents with each others. This allows for dynamic reconfiguration and continuous adaptation for optimal solutions. The agents can much better locally decide if a schedule for treatments is possible, impossible or borderline. Agents are in this context the mobile devices of all relavant paritcipants. An automatic replanning algorithm could for example first collect localization information of the patient and the next treatment‘s position. It will then consider the physical state of patients or delays of the treatment staff (for example because of meetings or emergencies). Together this information will often allow for dynamic re-organization of treatment schedules. The consequence is that many disturbances can be compensated by a little reorganization. The big benefit is, that this can be done “on-the-fly” without the need of a central coordination.

An example: assume patient A will not make it to his next treatment in time, while patient B, who is scheduled right after A, is already close to the treatment. It is then possible to inform A, B and the staff and to dynamically switch the patients‘ treatment. Note, that this simple scenario could also be possible without a computer aided planing system. But it is clear that the same method can also be used to make more complex re-schedulings, which may involve multiple changes. This is better than traditional approaches, because firstly re-planning starts as soon as a disturbance is detected and secondly it is not necessary for all agents to physically meet at one place but rather share and exchange information through the ad-hoc network. Similar algorithms can compen-

[1] A more detailed description of this example may be found in (Güdemann et al., 2006).

sate for delays of staff, broken equipment and many other disturbances. Furthermore the system could be used to *automatically* integrate new patients into the system by calculating treatment schedules for them. As a result of later automatic optimization some re-scheduling of existing schedules could make sense to improve the overall quality of the planning system.

All the properties described above are typical properties of Organic Computing systems. In Organic Computing these effects are called self-organizing (e.g. autonomously integrating new patients and calculating schedules), self-optimizing (e.g. minimizing waiting times), self-healing (e.g. compensating for broken equipment or illness of staff) and self-adapting (e.g. adapting to changing treatment requirements and health situation of a patient). So its only a logical step to apply techniques, tools and middlewares of Organic Computing to the domain of health care.

4 DOMAIN SPECIFIC CHALLENGES

What are the specific challenges to implementing an Organic Computing system in health care? The sad point is: traditional Organic Computing techniques are not directly applicable to this domain. The problem is, that most Organic Computing approaches make two assumptions of the systems and its components agents. The first assumption is, that all participants/agents share a common goal. The second one is, that there are no "malicious" parties involved. This means, all agents work for common interests and no agent tries to gain only benefits for himself while causing major drawbacks for all others. This arises a number of challenge, which must be mastered for successfully integrating Organic Computing in the domain of health care:

1) Individual Goals
Whenever agents of an Organic Computing system are personalized to individual persons, they in general won't share a common goal. In the example described above, agents owned by patients, medical staff and administration will have different goals. Patients are interested in treatment schedules, which are compatible with their leisure times, medical employees are interested in compact working times and the clinic's administration is interested in maximum profits i.e. no canceled treatments. These conflicting goals have to be balanced and weighted in a meaningful manner. It must also be assured, that security mechanisms prohibit individual persons from getting benefits on the cost of drawbacks for multiple others. So it will be a challenge for the next generation of Organic Computing algorithms to be able to cope with individual goals and requirements.

2) Sensitive Data
Personal data is sensitive. Medical data is even more sensitive. Therefore privacy and security is a prime requirement for every system, which handles data in a medical care scenario. On the other hand, the Organic Computing paradigm strongly relies on interaction and exchange of data between system components/agents. New and optimal configurations are computed at runtime jointly by all agents. This is only possible, because of information exchange and reflection of the exchanged data. For health care it will be a challenge to adapt Organic Computing algorithms such that only for reconfiguration and planning necessary data is exchanged and that sensitive data can be kept private wherever and whenever possible.

3) Behavioral Guarantees
Health care is a highly critical domain. Whenever a system is allowed to autonomously make decisions, which have effects on the medical treatment of a patient, it must be assured that these decisions may never pose a thread to the health of the patient. This problem can be solved by applying formal methods for analyzing the systems. They allow for rigorously proving, that the system will always fulfill some behavioral guarantees. For Organic Computing such analysis is much more difficult, because it can often not be anticipated at design time in which environments/scenarios the system will eventually be asked to (self-)adapt to. Therefore new analysis methods must be developed, which allow for giving behavioral guarantees for Organic Computing systems.

4) User Trust
This is possibly the most important challenge. Organic Computing systems can achieve a lot of benefits from a global point of view. The benefits result from their ability to self-organize and self-adapt. Improvements become typically visible on system wide performance metrics. For local agents, (re-) configuration is not necessarily beneficial or even traceable. This will also affect the user, who own the agents. If users are often confronted with decisions/plannings of the system, which they can not understand (and from which they don't benefit), then they will likely not accept the system. This must not happen. So it will be a challenge to provide algorithms, which provide users with enough information to understand and accept the decision of the system (without violating privacy constraints).

5 CONCLUSIONS

Organic Computing is a very promising new method for construction of modern software systems. It has proven to be an superior architecture in many domains like traffic control, sensor networks or production automation. Organic Computing is also very promising for a variety of user-intensive scenarios, where individual users own individual agents. The purpose of these agents can be to assist, support and/or guide their users. One scenario in health care is a planning/scheduling system in a rehabilitation clinic. An organic system can perform in this context much better than traditional approaches. On the other hand constraints, which arise in this domain – like concurrent/individual goals of users, privacy, behavioral guarantees and user-trust – require new Organic Computing algorithms and techniques. Developing such algorithms and methods can have a significant impact on many domains and open new opportunities and functionalities.

REFERENCES

Güdemann, M., Ortmeier, F., and Reif, W. (2007). Using deductive cause consequence analysis (DCCA) with scade. In *Proceedings of SAFECOMP 2007*. Springer LNCS. accepted at SAFECOMP07.

Güdemann, M., Ortmeier, F., and Reif, W. (2006). Formal modeling and verification of systems with self-x properties. In Yang, L. T., Jin, H., Ma, J., and Ungerer, T., editors, *Proceedings of the Third International Conference on Autonomic and Trusted Computing (ATC-06)*, volume 4158 of *Lecture Notes in Computer Science*, pages 38–47, Berlin/Heidelberg. Springer.

Güdemann, M., Ortmeier, F., and Reif, W. (2006). Safety and dependability analysis of self-adaptive systems. In *Proceedings of ISoLA 2006, 2nd Symposium on Leveraging Applications of Formal Methods, Verification and Validation*. IEEE CS Press.

Kephart, J. O. and Chess, D. M. (2003). The vision of autonomic computing. *Computer*, 36(1):41–50.

Müller-Schloer, C. (2004). Organic computing: on the feasibility of controlled emergence. In *CODES+ISSS '04*, NY, USA. ACM Press.

Müller-Schloer, C., von der Malsburg, C., and Würtz, R. P. (2004). Organic computing. *Informatik Spektrum*, 27(4):332–336.

Pietzowski, A., Satzger, B., Trumler, W., and Ungerer, T., editors (2006). *A Bio-Inspired Approach for Self-Protecting an Organic Middleware with Artificial Antibodies*, volume LNCS 4124. Passau, Germany.

Richter, U., Mnif, M., Branke, J., Müller-Schloer, C., and Schmeck, H. (2006). Towards a generic observer/controller architecture for organic computing. In Hochberger, C. and Liskowsky, R., editors, *INFORMATIK 2006 – Informatik für Menschen*, GI-Edition – Lecture Notes in Informatics, pages 112–119, Bonn, Germany. Köllen Verlag.

Seebach, H., Ortmeier, F., and Reif, W. (2007). Design and Construction of Organic Computing Systems. In *Proceedings of the IEEE Congress on Evolutionary Computation 2007*. IEEE Computer Society Press. accepted for publication.

Statistisches Bundesamt (2006). 11. koordinierte bevölkerungsvorausberechnung – annahmen und ergebnisse.

Statistisches Bundesamt (2007). Gesundheit – ausgaben 2005.

Trumler, W. (2006). *Organic Ubiquitous Middleware*. PhD thesis, Universität Augsburg, Eichleitnerstr. 30.

Trumler, W., Bagci, F., Petzold, J., and Ungerer, T. (September 20-24 2004). Towards an Organic Middleware for the Smart Doorplate Project. In *Workshop on Organic Computing, INFORMATIK 2004 - Informatik verbindet*, pages 626–630, Ulm, Germany.

VDE/ITG/GI (2003). Organic Computing: Computer- und Systemarchitektur im Jahr 2010. http://www.gi-ev.de/download/VDE-ITG-GI-Positionspapier Organic Computing.pdf.

USING WEB SERVICES TO DYNAMICALLY EMBED MEDICAL CONTENT IN A CLINICAL INFORMATION SYSTEM

Martin Luethi
Picis, Inc., 9550 W. Higgins Rd, Rosemont, IL, 60018, USA
martin_luethi@picis.com

Keywords: Web services, medical calculators, drug dosage calculators, dynamic user interface.

Abstract: This paper describes a prototype software application that makes use of Web services to retrieve medical content from an external Application Service Provider (ASP) service. The retrieved data is used by the system to dynamically generate user interface components within a user-customizable department information system and by clinicians to obtain results that will assist with medical decision-making. Medical forms are frequently built and rearranged by medical system administrators. From the perspective of clinical end users the content is seamlessly integrated and allows querying of dosages, interactions, and formulations using collected parameters such as age, weight, and physiological data. Whereas the use of standalone personal digital assistant (PDA) devices provides measurable benefits to clinicians, the feasibility of a seamless integration with user-customizable information systems has not been researched well. This paper describes one approach that could be taken to integrate such medical calculators with a web-based clinical information system. The solution described is not intended to represent functionality available in any actual or planned product.

1 INTRODUCTION

Medical content and knowledge such as drug interactions, medical calculators, drug dosing calculators, drug formulations, and illustrations provide a high benefit to clinicians and may decrease patient risk significantly (Knollmann et al., 2005). Medical knowledge has been made available in vast compendiums and in more recent days on the internet or as standalone applications on mobile devices such as PDAs. Whereas these applications provide a measurable benefit, they lack integration and documentation capabilities with clinical information systems that are concurrently used in the department. Consequently retrieved results have to be duplicated using the departmental procedures and systems in place.

Often departmental systems provide this medical content as well. It is either provided by the software vendor directly or external databases are purchased from vendors and then tightly integrated with the application. However, the conventional approach has two major disadvantages: First, changing content requires a new release or system maintenance cycle. Second, the integration is often static and does not allow the user to modify the clinical content.

The presented approach provides a solution to these problems. Using an ASP model, the content is directly queried from the content provider via Web services. Updates of data and newly added functionality are immediately available at sites where the system is in use. Further, the user interface (UI) is instantiated in an ad hoc fashion and the content is integrated within the dynamic structure. Updated content is immediately available for use and required UI elements are rendered. Patient data already available in the departmental system such as age, weight, and height is automatically pulled into the calculators. Once all necessary parameters are collected an additional Web service call is made to retrieve the calculated result based on the parameters provided by clinicians.

This paper also describes security challenges specific to the use of external Web services with medical data. Availability and reliability issues are discussed as well as general patient safety risks and testing strategies. Scenarios are discussed from the perspective of the service requester.

2 CLINICAL CONTENT

Medical knowledge is increasing rapidly. It is difficult for clinicians to keep up-to-date with new evidence-based studies, medical literature, and FDA-approved medications and their interactions, even though they should be considered during medical decision making. Prescription errors and adverse drug events are a leading cause for malpractice litigation (De Sousa, 1996). Improving access to this knowledge could lead to better decision-making, improve patient outcome, reduce costs, and increase bed utilization (Pronovost et al., 2002).

Portable devices, namely PDAs, have been used to make this knowledge accessible for clinicians (Lapinsky et al., 2004). The largest criticism of these systems besides data-related issues is their lack of integration with departmental information systems, their lacking ease of use, and the need to manually update the devices' databases. Content is preferably accessed in one single application and used on desktop computers when away from the bedside. Subsequently, examples are provided for often accessed content types.

2.1 Drug Dosage Calculators

Drug dosage calculators are used to calculate medication doses taking into account individual patient data such as weight and age and other relevant information. The applied formulas may vary by treated disease and type of medication.

The following table presents an example of a simple calculation formula (1) for IV Dobutamine. The parameters required for the calculation a-d are listed.

a = drug amount (mg)
b = per fluid (ml)
c = weight (kg)
d = dose (mg/kg/min)
x = drip (ml/min)

$$x = a / b * c * d \quad (1)$$

2.2 Medical Calculators

Medical calculators are used to generate a numeric or textual result that is used for further medical decision making. Typical examples are the calculation of probabilities of diseases and outcomes, index numbers, and scales. They can have one or multiple data entry parameters, which can be numeric or chosen from a selected textual option.

The following example is Basal E Expend for females (2), which requires weight, height, and age as input parameters and results the calorie use per day.

a = weight (kg)
b = height (cm)
c = years
x = kcal/day

$$x = 9.6 * a + 655 + 1.8 * b – 4.7 * c \quad (2)$$

2.3 Drug Interactions

The intake of more than one drug can result in severe interactions. Before prescribing or applying drugs, possible interactions have to be checked. With the current prototype several dozens can be checked simultaneously and the seriousness of interactions is rated on a scale of 1-5. Each interaction is accompanied with an explanation of the effects of the interaction.

As an example the interaction between Aspirin and Atenolol is provided. The result contains the severity of the interaction and its description. To obtain the desired result the drug identifiers have to be provided.

a = Aspirin (drug 1)
n = Atenolol (drug n)
x = drug interaction (level, message)

$$x = a, \ldots, n \quad (3)$$

2.4 Other Content

A simple case of medical content is the provision of textual and visual data regarding a contextual topic. The difficulty is to provide an appropriate search term to retrieve the data dynamically depending on the context and to link it with the application.

2.4.1 Text Content

Medical text content requires one or more search terms and a product code. Access of an encyclopaedic database is currently not supported. The service returns a result of drugs and products in textual representation.

2.4.2 Illustrations

Images are used for educational purposes. They provide explanations and textual values (Figure 1). Images are static and do not need any input parameter except the expected graphic format and the illustra-

tion identifier. The binary output is embedded in the XML body as base-64 encoded string.

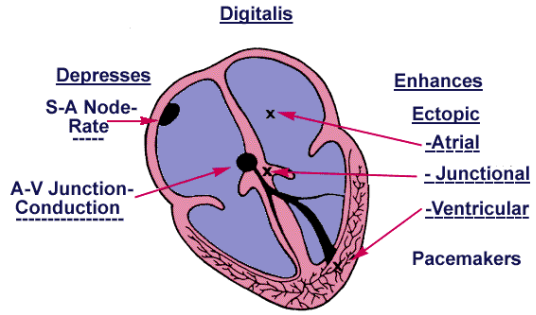


Figure 1: Example of a medical illustration.

3 WEB SERVICES

A Web service is a software system that is designed to support interoperable machine-to-machine interaction over a network. Its interface is described in a machine-processable format so that other systems can interact with the service exchanging messages (W3C 2004). For a successful message exchange mechanics and semantics of a message have to be defined and mutually agreed upon.

The mechanics describe the message format, data types, transport protocols, and transport serialization protocols. Commonly these definitions can be found in a Web services Description Language (WSDL) document. The semantics give the message meaning and its definitions are less standardized and not necessarily written or negotiated.

The use of Web services in a Service-Oriented Architecture (SOA) allows for the flexible and loose coupling of applications in platform and language independent manner. It accelerates application integration and facilitates rearranging the sequence of required components as needed. However, communications of distributed systems are naturally less reliable and slower than direct code invocation and shared memory. Further, there are additional security and privacy consideration in particular for services transmitting confidential patient information and cases in which the Web service data is directly used for patient treatment.

Application of Web services in clinical areas and related-fields as opposed to payer and business areas are not very common. A few examples of their use have been described (Cheng, Yang, Chen, Chen, & Lai, 2004; Eaton, 2006; Jiang, 2004; Mykkänen, Riekkinen, Sormunen, Karhunen, & Laitinen, 2007).

3.1 Standards

The Web services programming stack (Figure 2) is a collection of standardized protocols that are used to implement Web services (Gottschalk, Graham, Kreger, & Snell, 2002). The foundation of distributed Web services is the network layer. It is based on HTTP allowing accessing the service from within a hospital network through its firewalls. On top of the network layer, an Extensible Markup Language (XML)-based messaging protocol ensures the mechanics of the Web services. The Simple Object Access Protocol (SOAP) is used to invoke method calls for the content needed. Descriptions in WSDL give details of the mechanics and the available services of the interface. Available services and their location (URI) are known to the service requesters by informal means. However, publication and discovery of Web services could be implemented using the Universal Description Discover and Integration (UDDI) standard. The service flow layer allows the composition of services to an application. It has been implemented proprietarily.

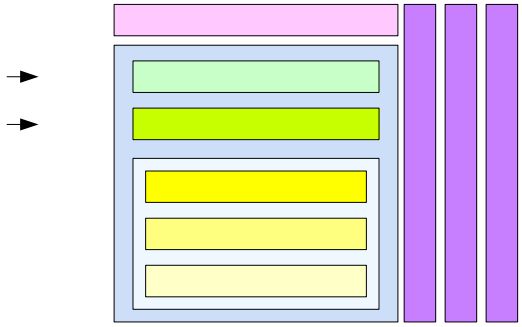


Figure 2: Web services programming stack.

Security, manageability, and quality of service apply to all the discussed layers.

3.2 Security and Reliability

Web services are suspect to common threats of networked data exchanges such as message alteration, confidentiality, man-in-the-middle, spoofing, denial of service, and replay attacks (W3C 2004). Such attacks can be prevented by point-to-point security technologies that secure the transmission from one Web services to another or, if routed through multiple services, by end-to-end security measures.

For the use of medical content as described, message alteration and denial of service are possible security risks. However, point-to-point technologies are sufficient to secure the transmission since messages are not routed among service providers. To

achieve this security, password authentication of the requester and encryption of the communications link using the Secure Socket Layer (SSL) protocol are considered sufficient. Confidentiality attacks and privacy concerns are not relevant for the discussed system since there is not enough data transmitted to identify a patient. Age and gender in conjunction with physiological data alone is in most cases not sufficient to deduce the identity of a patient.

Additionally, since the information system is used in an information technology (IT) environment that is not directly controlled by either party, it is important to consider that Web service calls need to be able to flow through organizations' firewalls. It can be tunnelled through existing ports and protocols, which makes this less of an issue.

Requesting services over a public network involving a requesting and providing agent, is intrinsically less reliable than local services. It can be distinguished between message reliability and service reliability. If no answer can be retrieved an error message is thrown. It would be possible to offer a fail-over service provider that could be hosted within a hospital network. Before values are submitted to the service, the values are checked on the client. They are checked for their validity and reviewed by the clinician when documenting a visit. Failures are not safety critical.

4 APPLICATION

The Web services consuming software application is a web-based emergency department information system (EDIS) that provides functionality required by clinicians such as patient tracking, registration, clinical workflow and task management, physician and nursing documentation, orders, and post-disposition management. The server-based application is interfaced with several other information systems in the hospital such as the main registration system, laboratory systems, pharmacy systems, billing and coding systems, etc.

The workflow in emergency departments depends on forms and documentation requirements vary significantly from site to site. Therefore, the clinical documentation component allows medical administrators to build and customize their own forms. The form building process may be performed via a drag-and-drop interface where UI elements are selected from a library of elements. Further, metadata such as charge codes, clinical identifiers, risk elements, and wording data can be attached. The content elements could be stored in a library.

Once customizations have been finished the content is published and is used for documentation by clinicians. Documentation is complaint-driven and can consist of many different forms, whose sections are selected dependent on patient attributes such as age and gender.

4.1 Architecture

Clinical documentation forms are created in the content builder component. The elements that make up a form are stored in a relational database containing the metadata, but no presentation layer data. The data is separated from the presentation layer, which allows using the form definition data in a different context such as for PDAs and paper forms for digital pen devices.

Figure 3 illustrates the process of using the Web service. In the content builder component the clinical form is assembled. At the location where a medical or dosage calculation is required, a generic calculator widget is positioned. A list of available calculators is presented to the user. This list could be retrieved through the Web service also, but is kept locally for performance and manageability reasons. The current prototype can contain hundreds of different calculators.

When the calculator is selected a Web service call is made to the ASP service with the needed calculator identifier (1). The Web service responds with an XML document containing the field data for the requested calculator. In the response, each required data field is specified with label, variable name, and data type. If the data type is an enumeration all possible options including their text and data representation are included. Further the formula used to calculate the result is provided in Reverse Polish Notation (RPN), where applicable.

The response is parsed and the UI elements are dynamically generated (2). Since the presentation layer is built using an object-oriented hierarchy of classes the received data has to be mapped so that the necessary objects can be instantiated. The entire form containing also non-Web service retrieved elements is then rendered to be used in the application. During form creation, response time and performance are not a primary concern since the form is compiled into an intermediary format and cached locally.

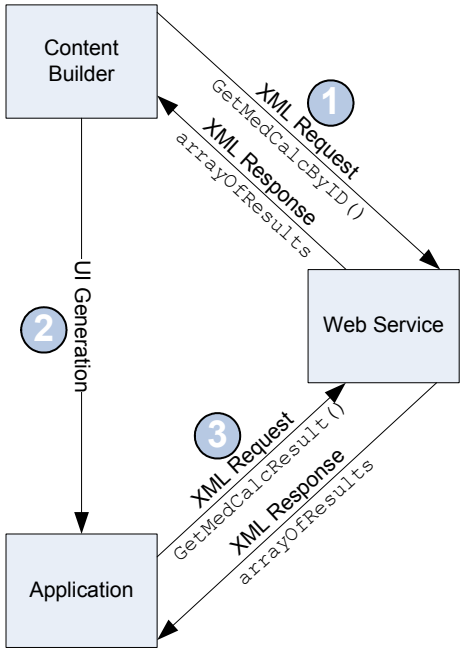


Figure 3: Architecture.

The response is parsed and the UI elements are dynamically generated (2). Since the presentation layer is built using an object-oriented hierarchy of classes the received data has to be mapped so that the necessary objects can be instantiated. The entire form containing also non-Web service retrieved elements is then rendered to be used in the application. During form creation, response time and performance are not a primary concern since the form is compiled into an intermediary format and cached locally.

During system use by clinicians the required parameters are filled in and an Asynchronous JavaScript and XML (Ajax) request is made to the Web service to calculate the results (3). The score/result field is populated with the result as soon as available. The value can be a string or numeric type depending on the calculator. Due to the asynchronous nature of the call, the response does not have to be instantaneous and some delays are tolerable. If the service is unavailable or does not provide the expected result an error message is thrown. Alternatively, the request could be routed to a different service, handled locally, or queued for later retrieval.

4.2 Visualization

The XML response of the Web service containing the calculator data looks as presented below. Additionally there is a WSDL service description available that describes the expected response.

```
<MedCalcInfo …>
  <Error />
  <MedCalcID>bish</MedCalcID>
  <Title>Bishop</Title>
  <Inputs>
    <Input>
      <Name>Dilation</Name>
      <Var>a</Var>
      <InputType>Enum</InputType>
      …
      <Options>
        <Option>
          <Text>Cervix … <1 cm</Text>
          <Value>0</Value>
          <Selected>true</Selected>
        </Option>
        …
      </Options>
    </Input>
    …
  </Inputs>
</MedCalcInfo>
```

The response is parsed and the object hierarchy is instantiated. The response also specifies the default values for enumerations as well as data types that are required. The rendered interface enforces data type checking using client-side validation. The UI also consists of data types unknown to the Web service. For example, in medicine the concept of pertinent negatives is constantly encountered. This means that such an element can assume three different states: not answered, positive, or negative. In the UI this concept is implemented as a checkbox that can be negated by cycling through its different states by clicking the mouse repeatedly. Positive values are represented in a circled manner; negative values are displayed slashed. This application specific UI component could be considered a type of “SuperCheckbox”.

Figure 4 shows how the different data types are displayed for the end users. This concept has to be mapped from the Web service response. Enumerations that are specified as yes/no option value are visualized as SuperCheckbox instead of a dropdown.

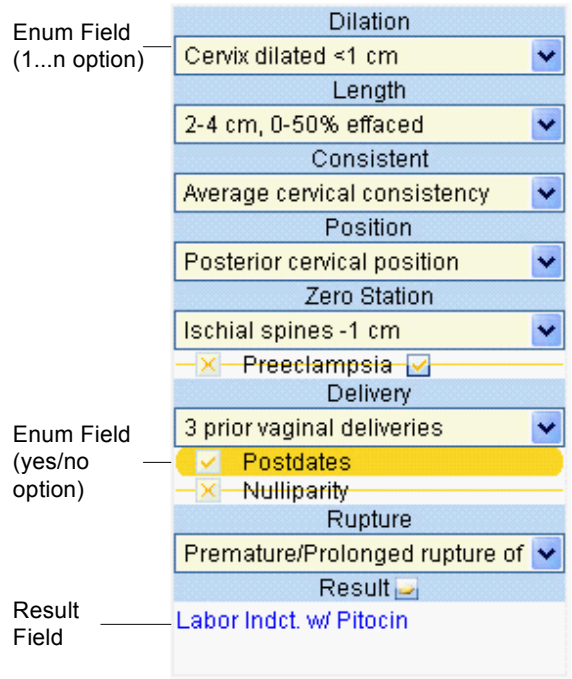


Figure 4: Dynamically generated user interface.

The UI element types are selected according to the criteria presented in Figure 5.

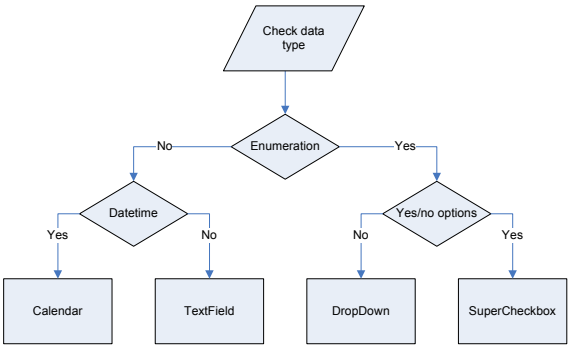


Figure 5: Data type mapping.

Additional to the field definitions received from the Web service, a result field is rendered to present the calculated scores or results. The results and values do not only have to be submitted to the Web service for calculation but also documented in the EDIS according to site-specific guidelines. The request and response are validated by the service whether they lie in a sensible range.

4.3 Integration

Many values that are used in calculators are already available in the system either through manual data entry using different forms or through data interfaces with external systems. Such values include vital signs, weight, height, age, gender, etc. These values can be automatically imported and pre-selected in the appropriate UI elements. To provide real semantic integration, a system would require the use of a common nomenclature. String parsing, mapping, and data type checking could be applied, which in conjunction would provide the required reliability.

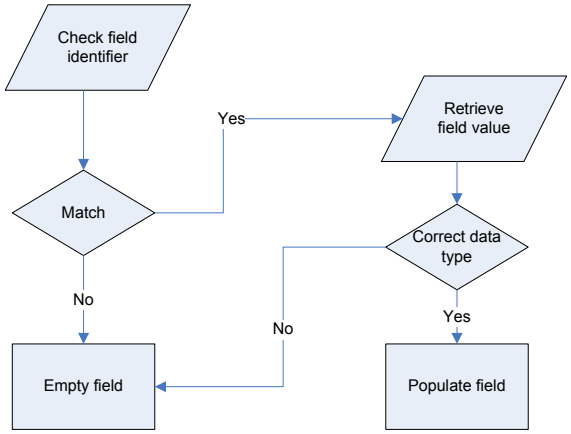


Figure 6: Populate fields with existing data.

The quality of the described system can be assured by automated unit tests that query the Web services for several test cases. They also record the access time and latency of the service response time. The UI is also tested via a test automation framework.

4.4 Measurements

As indicated above, response times and reliability of the discussed Web services are of some concern. This is in particular the case for the calculation of medical results. Whereas some delays are acceptable due to asynchronous calls, response times should generally not be beyond a 4 second threshold.

The measurements were taken querying the Web services over the internet over a distance of 14 hops. The average roundtrip time was 43 milliseconds. The services were tested with 1 to 20 concurrent clients. For each client 100 samples were taken. The infrastructure is scalable but further prototype development is required to identify practical limits.

Table 1 shows the results for retrieving the calculator data during form administration. During the test, all requests could be completed successfully. Generally not more than 5 concurrent users are expected to use the administration tool at any given time. Rendering times of the data are negligible.

Table 1: Response times in seconds of calculator service.

Clients	Average	Minimum	Maximum	Deviation
1	0.4930	0.3750	0.7500	0.0638
5	1.8393	0.4220	2.5310	1.1898
10	3.7522	0.4840	12.3140	1.8493
20	8.8832	0.3590	62.3220	5.7396

Table 2 lists the response times of the Web service that calculates the results. It performs slightly better using the same number of concurrent clients. An acceptable wait time is arbitrarily set at 4 seconds. To process more concurrent clients beneath that threshold the infrastructure would need to be load-balanced.

Table 2: Response times in seconds of result service.

Clients	Average	Minimum	Maximum	Deviation
1	0.6528	0.3120	3.0160	0.3870
5	3.1357	0.3440	14.1410	3.8315
10	3.4759	0.4530	12.9840	1.3377
20	7.8717	0.6250	27.4390	3.3215

Table 3: Conclusions.

Advantages
Immediate updates of frequently changing medication information and no maintenance of additional servers.
Vast amount of available medical content for immediate use with minimal integration effort.
Seamless integration with clinical documentation.
Ability of user customization of clinical content.
Disadvantages
Dependence on external service.
Potential service outages and response time problems.
Regulatory and liability implications need to be determined.

5 CONCLUSIONS

This paper showed how medical content can be dynamically integrated into a clinical information system. The primary challenge was the seamless integration of external content into a customizable UI. The components are rendered on-demand but still can be transparently arranged by medical system administrators.

Whereas the retrieval of the form definition data is only performed during system customization, the retrieval of results occurs during the documentation by clinicians and requires a higher level of availability. Reliability was found to be high for the described service and security requirements can be fulfilled by point-to-point security technologies.

It was found that the advantages generally outweigh the disadvantages. The potential of having a vast amount of external content available without a significant integration effort was considered to be the biggest advantage. Further, having this content available using an ASP concept as opposed to in-house hosting reduces maintenance to a minimum for the service requester.

Future research is needed to address the acceptance of this content by clinicians. Also, the response times and reliability over a high number of customers needs to be considered. In a next step, patient history from other systems could be integrated using Web services across network boundaries. Also, legal and regulatory concerns need to be addressed.

REFERENCES

Cheng, P., Yang, C., Chen, H., Chen, S., & Lai, J. (2004). *Application of HL7 in a collaborative healthcare information system.* Paper presented at the Engineering in Medicine and Biology Society, 26th Annual International Conference of the IEEE.

De Sousa, B. (1996). Prescription problems: clinical pharmacology and lawsuits. *Texas Medicine, 92*(5), 57.

Eaton, A. D. (2006). HubMed: a web-based biomedical literature search interface. *Nucleic Acids Research, 34, W745–W747*.

Gottschalk, K., Graham, S., Kreger, H., & Snell, J. (2002). Introduction to web services architecture. *IBM Systems Journal 41(2), 170-177*.

Jiang, K. (2004). *Integrating clinical trial data for decision making via web services.* Paper presented at the Engineering in Medicine and Biology Society, 26th Annual International Conference of the IEEE.

Knollmann, B., Smyth, B., Garnett, C., Salesiotis, A., Gvozdjan, D., Berry, N., Lee, H., & Min, D. (2005). Personal digital assistant-based drug reference software as tools to improve rational prescribing: Benchmark criteria and performance. *Clinical Pharmacology & Therapeutics, 78*(1), 7-18.

Lapinsky, S. E., Wax, R., Showalter, R., Martinez-Motta, J. C., Hallett, D., Mehta, S., et al. (2004). Prospective evaluation of an internet-linked handheld computer critical care knowledge access system. *Critical Care, 8*(6), 414-421.

Mykkänen, J., Riekkinen, A., Sormunen, M., Karhunen, H., & Laitinen, P. (2007). Designing web services in health information systems: from process to application level. *International Journal of Medical Informatics, 76*(2-3), 89-95.

Pronovost, P. J., Angus, D. C., Dorman, T., Robinson, K. A., Dremsizov, T. T., & Young, T. L. (2002). Physician staffing patterns and clinical outcomes in critically ill patients. *The Journal of the American Medical Association, 288*(17), 2151-2162.

W3C (2004). *Web services architecture*. Retrieved 2007-06-11, 2007, from http://www.w3.org/TR/ws-arch/wsa.pdf

SMART TRANSPLANTATION

Fever of Unknown Origin after Stem Cell Transplantation as a Model for a Knowledge-Based Decision Support System in Medicine

Gerrit Meixner[1], Nancy Thiels[2]
[1]*University of Kaiserslautern, Gottlieb-Daimler-Strasse, Kaiserslautern, Germany*
[2]*German Research Center for Artificial Intelligence (DFKI), Trippstadter Strasse, Kaiserslautern, Germany*
[1]*meixner@mv.uni-kl.de,* [2]*nancy.thiels@dfki.de*

Ingo Haschler[3], Andreas Wicht[4] and Ulrike Klein[5]
University of Heidelberg, Im Neuenheimer Feld, Heidelberg, Germany
[3]*ingo.haschler@med.uni-heidelberg.de,* [4]*email@andreas-wicht.de,* [5]*Ulrike.Klein2@med.uni-heidelberg.de*

Keywords: Human-Computer Interaction in Health Care, Usability of Medical Information Systems, Human Aspects of Future Technologies in Health Care, Cognitive Task Analysis, Usability Engineering, Stem cell transplantation, Decision support system.

Abstract: Public health care has to make use of the potentials of IT to meet the enormous demands on patient management in the future. Embedding artificial intelligence in medicine may lead to an increase in health care quality and patient safety. One possibility in this respect is the use of knowledge-based decision support systems which facilitate the practice of evidence-based medicine. Conditions for such a system are structured data sources to extract relevant data for the proposed decision. Therefore, the demonstrator “allo-tool” was designed. To develop the allo-tool a user-orientated process was applied and future users of the later software were integrated in each step of the development process. The concept of introducing a "Medical decision support system based on the model of Stem Cell Transplantation" was developed afterwards. The global objectives of the planned system are (1) to improve patient safety (2) to support patient autonomy and (3) to optimize the work flow of medical personnel.

1 INTRODUCTION

In many areas of human life, computer-based Information Technology (IT) has prevailed and has become essential for the coordinated and efficient organization of work flow. This offers numerous advantages for the future, but will also lead to problems for the people confronted with it. Especially in the field of health care, interaction between human beings and information technology is a sensitive subject. Physicians have immense reservations and apprehensions of being made the slaves of information scientists and of their programmed computer system. Nevertheless, medicine has to become more scientific in patient management. The importance of interdisciplinary advice and discussion which is a prerequisite for the best possible decision on a treatment strategy has to be reflected in the application of new information technologies.

Embedding artificial intelligence in medicine may lead to an increase in quality and safety and to a decrease in costs significantly. In the future, IT-systems are supposed to have the ability to extract the relevant knowledge to filter irrelevant information and focus on significant information and to present it available to the user. Inevitably, people working in health care will have to make use of the potentials of IT in order to meet the enormous demands on patient management in the future.

One possibility in this respect is a knowledge based decision support system which facilitates the practice of evidence-based medicine. The idea of medical decision support systems is not new. In the past, however, these systems have not become popular and utilized. The reason was not that the technology had failed, but that the implementation was inadequate. A prerequisite for success is that we understand "how medicine thinks" in order to be

able to create a "decision-supporting", and not a "decision-making" system.

Decision support systems need an evaluation of their applicability and of their cost-benefit relationship as well as the evaluation of precision, quality, and effects of their recommendations. Against the background of the challenges mentioned above, the concept of introducing a "Medical decision support system based on the model of Stem Cell Transplantation", named allo-tool, was developed.

Whenever bringing a decision support system into clinical practice, one has to consider the possible harms caused by the system. Concretely this means to consider what happens if the system recommends a wrong decision or presents incomplete or inadequate information at the point of care to the physician. We therefore plan to integrate a feedback mechanism. In the case that a physician has the opinion that the system could recommend something wrong, he will submit an incident report to the maintenance team. In that way, the users (i.e. mainly the physicians) will retain control over the system and possible harm to the patients caused by the system will be minimized.

The global objectives of the planned system are (1) to improve patient safety (2) to support patient autonomy and (3) to optimize the work flow of medical personnel. It might lead to more efficient use of resources without detrimental effects on the relationship between physician and patient or on the physician's autonomy to decide. The allogeneic haematopoietic stem cell transplantation is extremely well suited as a model for this type of system because of repeating standard procedures, the well defined span of time required and the predictable recovery period as well as recurrent side effects after transplantation.

2 STEM CELL TRANSPLANTATION

Hematopoietic stem cell transplantation (HSCT) or bone marrow transplantation (BMT) is a medical procedure in the field of hematology and oncology (Korbling, 1986). It is most often performed for people with diseases of blood or bone marrow (Gratwohl, 2007; Goldschmidt, 2000). HSCT remains a risky procedure and has always been reserved for patients with life threatening diseases. Since the availability of stem cell growth factors, most hematopoietic stem cell transplantation procedures have been performed with stem cells collected from the peripheral blood (Montgomery, 2007). Most recipients of HSCTs are patients with leukemia or aggressive hematological tumors who would benefit from treatment with high doses of chemotherapy or total body irradiation. Other patients who receive bone marrow transplants include pediatric cases where patients have an inborn defect and were born with defective stem cells. Other conditions that bone marrow transplants are considered for include inherently diseases of the bone marrow. More recently non-myeloablative, or so-called "mini transplant," procedures have been developed which do not require such large doses of chemotherapy and radiation (Djulbegovic, 2003). This has allowed HSCT to be conducted in older patients and as a matter of principle without the need for hospitalization. There are two major types of stem cell transplantation maneuvers: Autologous HSCT involves isolation of HSC from a patient, storage of the stem cells in a freezer, high-dose chemotherapy to eradicate the malignant cell population at the cost of also eliminating the patient's bone marrow stem cells, then return of the patient's own stored stem cells to their body. Autologous transplants have the advantage of a lower risk of graft rejection, infection and graft-versus-host disease. Allogeneic HSCT, as the second type, involves two people, one is the healthy donor and one is the recipient. Allogeneic HSC donors must have a tissue type that matches the recipient and, in addition, the recipient requires immunosuppressive medications. Allogeneic transplant donors may be related or unrelated volunteers.

The number of performed HSCT, autologous and allogeneic, is increasing. Due to better anti-infective medication the life-threatening side effects of infectious complications are decreasing but they still remain as the main risk factor for life threatening side effects (Walker, 2007; Afessa, 2006).

Therefore one intention of the allo-tool is the improvement of patient safety through a decision support system for choosing an anti-infective therapy based on evidence based advices.

Bacterial, viral or fungal infections are severe side effects of high dose chemotherapy and stem cell transplantation and can result in life threatening complications. Infections are the most important causes of morbidity and mortality in patients undergoing allogeneic stem cell transplantation. That is why immediate anti-infective therapy in case of fever is mandatory (Afessa, 2006). Because of the resistance to widely used antibiotics and a shift of

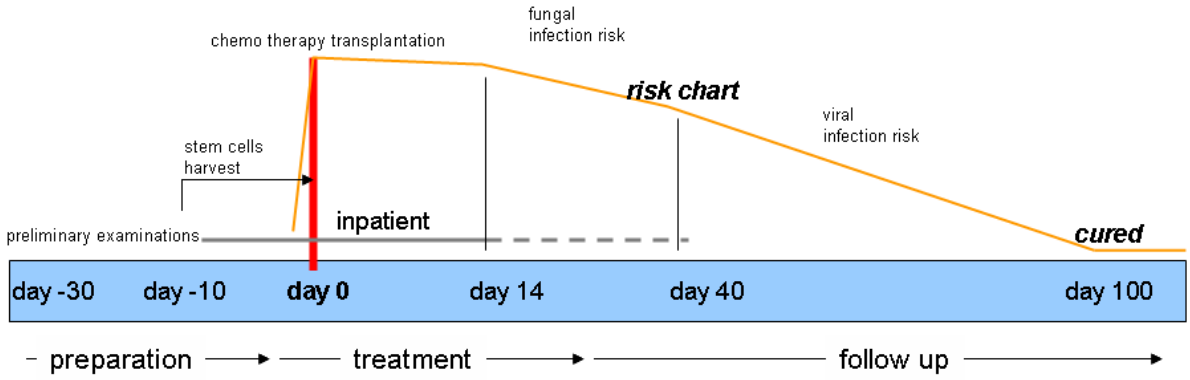


Figure 1: Risk chart graphic.

causative pathogens towards multi-resistant bacteria, anti-infective therapy remains to be a challenge. Beside this drug-interaction, kidney function, known allergies, the cause of fever and former anti-infective therapies needs to be included into the decision of the treatment of choice. Therefore the model of a decision support system was developed. The idea of programming and implementing such a system is not new. Attempts of this type had already been initiated in the Eighties (then called medical expert systems) (Shortliffe, 1975). In the past, however, these medical expert systems have not become popular and utilized. The reason was not that the technology had failed, but that the implementation was inadequate.

Beside this the support of patient autonomy through modern forms of communications like web-based access to specific patient data and the optimization of the work flow in the complex process of an allogeneic transplantation is the goal of the allogeneic tool “smartTransplantation”.

3 USEWARE ENGINEERING PROCESS

The level of acceptance and efficiency of a modern user interface are determined by the ease of use of the interface. Primary considerations in this evolutionary development process are always the requirements and needs of the user, for whom the user interface is being developed. The process consists of analysis, structural design, rough and detailed design, implementation, and evaluation (Zühlke, 2004). As a continuation of the analysis, evaluation occurs concomitantly to the development.

A risk chart (see Fig. 1) has been developed to get an understanding of the transplantation procedure. It explains the work flow of the stem cell transplantation in detail. The preparation of the analysis phase derived five different user groups. From these user groups ten different persons were chosen to be questioned to find out the tasks and needs of users for the later user interface.

After the questioning and characterizing, the collected data were compiled (Bödcher, 2007). Additional results included an analysis of weaknesses identified in the existing user interfaces and documentation of the “wish list” expressed by the users. Problems to be found within the analysis of the clinical situation for the stem cell transplantation were published.

Based on the task models of the different user groups, the use structure for the future user interface was developed. The modeling language *useML* (Reuther, 2003) was used in structural design. The result of the structural design phase is a platform-independent model and it provides the foundation for the later design phase. This model formed the base for the development of the prototype.

4 PROTOTYPE: ALLO-TOOL

The intention of the allo-tool is not only the optimization of the work flow in the complex process of an allogeneic transplantation, but also the provision of a structured platform for relevant data. This aim is reached by data integration from different existing information sources: clinical information system, drug information system and paper patient files as well as domain knowledge formalized in knowledge bases (described in more detail in chapter 5). The digitized aggregated data is displayed in a clearly structured way.

The tool shall be able to extract medical information from different sources, structure the

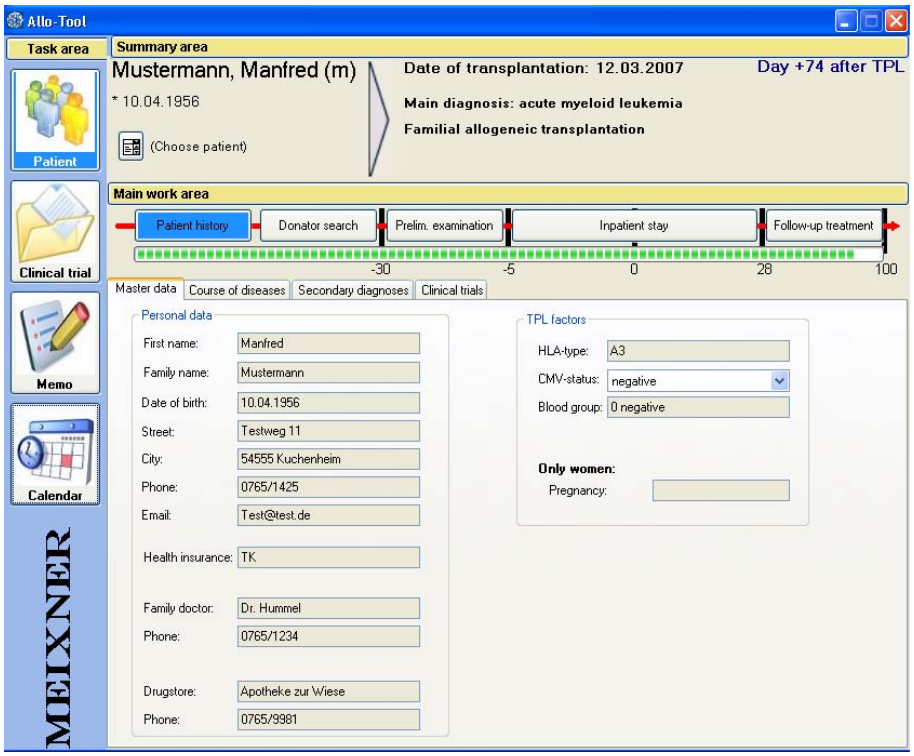


Figure 2: Patient history screen in the detailed view.

information and automatically generate discharge letters and further documents, e.g. drug plans.

Each user group is able to work consistently with the allo-tool. The tool consists of four main views: calendar, memos, clinical trials and patients. The memo view consists of a text field, where the user has the possibility to take notes. In the clinical trial view the user is able to administrate information about current clinical trials.

For allogeneic transplantations information about conditioning, transplantation, immunosuppression, GVHD prevention (graft versus host disease) and donor lymphocyte transfusions are stated within these clinical trials.

According to the work flow of an allogeneic transplantation, the detailed patients view consists of five different chronological parts in the main work area of the tool: patient history, donator-search, preliminary examinations, inpatient stay and follow-up treatment. These five parts are clearly separated phases in the time line of an allogeneic transplantation (see Fig. 1). The graphical representation of a time line (see Fig. 2) enables physicians to have a better understanding (Norman, 1990) of the current date according to the complex process of an allogeneic transplantation. The tool supports physicians in keeping deadlines according to a physical examination time schedule.

The patient's history phase (see Fig. 2) consists of a structured overview of the patient's master data. In the course of diseases previous examinations and medical results are displayed in a table view. Secondary diagnoses are an important part of the decision supporting system and help to minimize adverse effects.

The preliminary examination phase contains information about the accomplished examinations according to the physical examination time schedule. The results of external examinations are entered into the clinical information system. The allo-tool shall be able to extract the results of external examinations.

In the phase of hospitalization, the patient receives high dose chemotherapy and some patients a whole body irradiation. After that the allogeneic stem cell transplantation is performed. During the whole phase patient data (e.g. vital signs, blood counts or organ functions) are monitored very closely. Beside others, one important task of the attending physician is to administrate drug plans.

Via an implemented interface, physicians are able to use an existing drug information system (Kaltschmidt, 2004; Pruszydlo, 2006).

In the follow-up phase the physician is reminded to accomplish examinations according to the physical examination time schedule. The time schedule consists of a large number of physical examinations and complex execution logic. The prototype transfers the execution logic from the physicians to the software.

Patients in the follow-up phase have the possibility to access test results via internet and can directly communicate with their allocated physician via email. This will be reached by developing a web-based access for patients. Patients will have the possibility to view their own data (e.g. blood test results, X-ray photographs). Test results, which can be viewed by patients over internet are evaluated and released by their allocated physician. Patients in the follow-up phase have to control e.g. blood pressure values. They can measure blood pressure on their own and email the results over the web-based access directly to their allocated physician. So patients don't have to call the hospital via telephone or have to visit the hospital on their own.

5 KNOWLEDGE-BASED SUPPORT

In the future, physicians will be supported by a knowledge-based system with data mining capabilities. The detailed functionality is explained in the following.

Monitoring and interpretation of several parameters such as vital signs or laboratory results will be one main function of the knowledge-based system component. The system thus detects specific clinical situations and pushes unsolicited warnings or reminder and starts the according work flow.

In everyday clinical practice physicians are often faced with an information overflow rather than a lack of information, they have to spend valuable time in looking for relevant findings. To reduce this time a "Semantic Information Extraction" is provided. The system looks for findings in the patient's history which could fit into the context of the current clinical picture. Furthermore, it visualizes the classification of the findings, which has previously to be provided by a physician: "abnormal", "no abnormality", "unclear". So the physician gets a clearly arranged listing of relevant findings matching the current issue.

For a defined amount of clinical pictures, therapy and diagnostic recommendations are provided. These recommendations are on the one hand based on so-called domain knowledge which is patient-independent knowledge on e.g. diseases or processes. Domain knowledge will mainly derived from clinical practice guidelines (CPG) or in-house standard operating procedures (SOP). On the other hand they are based on available patient data such as current parameters as well as the patient's general data (if any parameters are missing or out of date, the system will ask the user to enter these), the system should give exact medication and dosing recommendations. Once the system has generated a recommendation, it will also provide the user with a reasonable explanation. In this context it may also be helpful that the user is guided to external knowledge resources, which match the current clinical picture, such as local SOP-documents, relevant study protocols or other medical knowledge bases, if further information is requested.

Recommending a (drug) therapy at first requires a knowledge base which encapsulates the SOPs and the guidelines' knowledge. The process of formalizing guidelines and SOPs is a challenging and time-consuming task (Kaiser, 2007). We will try to build a semi-automatic, peer-reviewed process, i.e. the system will try to import the unstructured material and recognize as much structure as possible. In the next step a specially trained medical professional will then review and complete structural or semantic issues. Afterwards, another medical professional will review the work of the first one to ensure high quality of the formalized knowledge.

Once the therapy decision was accepted and started, the system supports the medical staff by monitoring not only the over-all treatment work flow, but also the flow of the specific therapy. Concerning the over-all treatment work flow, the system should give reminders or warnings if scheduled diagnostic procedures (e.g. ECG) are not done or assigned yet or scheduled medications (e.g. antibiotic prophylaxis) are not ordered. Regarding the therapy monitoring, the system should detect if a patient does not respond to the treatment (e.g. patient remains febrile) and recommend necessary steps and/or alternative therapies as well as indicated diagnostic procedures. On the other hand, the system should detect if a therapy was apparently successful and recommend further steps (e.g. to stop antibiotics and to start a prophylaxis again). In this context the system presents - as already mentioned before - reasonable explanations and guides the user to external knowledge resources if requested.

As medical knowledge is subject to ongoing changes, the knowledge base has to be maintained regularly. Changes can be triggered either by external factors (e.g. update of a guideline) or by internal factors. We therefore need a user feedback mechanism which enables a physician to mark a system's decision as potentially wrong. The user

feedback has to be analyzed and the need to change the knowledge base has to be assessed. All changes to the knowledge base must be carried out within a "bullet-proof" process like the one described above.

A critical issue when it comes to the implementation of decision support systems is user compliance. In our environment we face mainly physicians as users, but also patients in an advanced state of the software (e.g. test results via web-based access). We have to ensure usability of the tool (as described earlier) and ensure that physicians get a benefit from it. The patients are most likely to do everything to improve their therapy outcome because of their severe illness.

One reason why decision support systems often do not prevail in clinical practice is poor work flow integration (Bates, 2003). Since our software will cover the SCT-treatment process as a whole we can map the practical work flow. Physicians will more likely use a decision support system if they see a clear benefit from it. This means mainly time-saving as well as convenient access to all relevant information.

Because the decision support system will be seamlessly embedded as a component into the allo-tool, we can reach an optimal solution to this obstacle. All of the knowledge-based features we presented above will appear within the current work flow context. We provide the medical user with only the information he or she needs at a given clinical situation. Therefore we can expect a reasonable time saving for the users. By ensuring that we do not miss relevant information on the other side, we may raise quality of treatment and patient safety.

6 CONCLUSIONS

It is incontestable that people working in health care will have to make use of the potentials of IT in order to meet the enormous demands on patient management in the future. Beside this the quality of work can be supported by intelligent software which is able to extract, rate and provide the user with relevant data. Not least patients require more autonomy of their own health information data. To meet this challenges the demonstrator of the allo-tool was developed.

Time consuming data search, redundant information and vast numbers of needed software applications are reduced by displaying all data in one tool. User interfaces, which are designed in close relationship to known software products, developed with the support of different users during the whole process and consulting of usability experts facilitate an easy-to-use application. Time schedules, reminder of deadlines and coherent information about study procedures enable medical staff to work efficiently. Taking these analysis results as a basis, conditions for a medical decision support system are accomplished. In order to meet the exploding number of scientific perception, decision support systems are needed in the future to maintain the quality of medical decisions.

Through web-based access to selected health information, patients obtain more autonomy and responsibility. Summarizing the potentialities of the planned allo-tool, the goals mentioned at the beginning, (1) improvement of patient safety (2) support of patient autonomy and (3) optimizing the work flow of medical personnel are illustrated. Studies to evaluate these potentialities are needed to prove these advantages of the allo-tool.

ACKNOWLEDGEMENTS

This work was supported by the Gottlieb Daimler- and Karl Benz-Foundation.

REFERENCES

Afessa, B.; Peters, S.G., 2006. *Major complications following hematopoietic stem cell transplantation,* Semin Respir Crit Care Med, 2006, 27(3), 297-309.

Bates, D. W.; Kuperman, G. J.; Wang, S.; Gandhi, T.; Kittler, A.; Volk, L.; Spurr, C.; Khorasani, R.; Tanasijevic, M.; Middleton, B., 2003. *Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality,* J Am Med Inform Assoc, 2003, 10, 523-30.

Bödcher, A., 2007. *Methodische Nutzungskontext-Analyse als Grundlage eines strukturierten USEWARE-Engineering-Prozesses,* Fortschritt-Berichte pak, no. 14, Kaiserslautern: University of Kaiserslautern.

Djulbegovic, B.; Seidenfeld, J.; Bonnell, C.; Kumar, A., 2003. *Nonmyeloablative allogeneic stem-cell transplantation for hematologic malignancies: a systematic review,* Cancer Control, 2003, 10(1), 17-41.

Goldschmidt, H.; Egerer, G.; Ho, A.D., 2000. *Autologous and allogeneic stem cell transplantation in multiple myeloma,* Bone Marrow Transplant, 2000; 25, Suppl. 2, 25-26.

Gratwohl, A.; Baldomero, H.; Frauendorfer, K.; Urbano-Ispizua, A.; Niederwieser, D., 2007. *Results of the EBMT activity survey 2005 on haematopoietic stem cell transplantation: focus on increasing use of unrelated donors,* Bone Marrow Transplant, 2007, 39(2), 71-87.

Kaiser, K.; Akkaya, C.; Miksch, S., 2007. *How can information extraction ease formalizing treatment processes in clinical practice guidelines? A method and its evaluation,* Artif Intell Med, 2007, 39, 151-163.

Kaltschmidt J.; Gallin S.; Haefeli W.E., 2004. *Essential functional requirements for an effective electronic drug information system in a hospital,* Int J Clin Pharmacol Ther, 2004, 42:615.

Korbling, M.; Dorken, B.; Ho A.D.; Pezzutto, A.; Hunstein, W.; Fliedner, T.M., 1986. *Autologous transplantation of blood-derived hemopoietic stem cells after myeloablative therapy in a patient with Burkitt's lymphoma blood,* J AM Society of Hematology, 1986, 67(2), 529-532.

Montgomery, M.; Cottler-Fox, M., 2007. *Mobilization and collection of autologous hematopoietic progenitor/stem cells,* Clin Adv Hematol Oncol, 2007, 5(2), 127-136.

Norman, D., 1990. *The Design of Everyday Things,* Currency.

Pruszydlo, M.G.; Kaltschmidt, J.; Schmitt, S.P.W.; Markmann, A.; Bertsche, T.; Haefeli W.E., 2004. *Successful implementation of computer-aided prescribing in a hospital providing tertiary care,* Int J Clin Pharmacol Ther, 2006, 44:529.

Reuther, A., 2003. *useML – Systematische Entwicklung von Maschinenbediensystemen mit XML,* Fortschritt-Berichte pak, no. 8, Kaiserslautern: University of Kaiserslautern.

Shortliffe, E. H.; Davis, R.; Axline, S. G.; Buchanan, B. G.; Green, C. C.; Cohen, S. N., 1975. *Computer based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the MYCIN system,* Comput Biomed Res, 1975, Aug, 8(4), 303-320.

Walker, C.M.; van Burik, J.A.; De For, T.E.; Weisdorf, D.J., 2007. *Cytomegalovirus infection after allogeneic transplantation: comparison of cord blood with peripheral blood and marrow graft sources,* Biol Blood Marrow Transplant, 2007, 13(9), 1106-1115.

Zühlke, D., 2004. *Useware-Engineering für technische Systeme,* Springer.

CONNECTING HOSPITALS AND HEALTH CENTRES COULD BENEFIT MOST OF PORTO'S CHILDREN POPULATION

Current Trends in Paedriatic Patients' Mobility between Institutions Requires Implementation of Electronic Patient Records

Ana Edral[1], Ana Sofia Matos[1], Beatriz Soares[1], Carlos Leite[1], Joana Almeida[1], João Girão[1], João Nunes[1], Maria Isabel Pereira[1], Miguel Lobato[1], Mónica Silva[1], Rita Boaventura[1], Sofia Faria[1] Ricardo Cruz-Correia[1,2]

[1]*Faculdade de Medicina,Universidade do Porto, Alameda Prof. Hernâni Monteiro, Porto, Portugal*

[2]*Center for Research in Health Information Systems and Technologies*

med06161@med.up.pt, med06183@med.up.pt, med06075@med.up.pt, med06205@med.up.pt, med06033@med.up.pt med06227@med.up.pt, med06041@med.up.pt, med06081@med.up.pt, med06054@med.up.pt, med06096@med.up.pt med06117@med.up.pt, med06139@med.up.pt, rcorreia@med.up.pt

Keywords: Computerized patient record, medical record linkage, paediatrics, health institutions, Multi Institutional Systems.

Abstract: Nowadays, data is spread across many institutions. Aim: This study aims to establish the need for the implementation of a regional Electronic Patient Record (EPR) in Porto area, to support the mobility, and the characteristics of the population that attends Porto's Paediatric Unit (UPP) Methods: The study is crossectional with a consecutive sampling method. A questionnaire was applied during three days. The study population consisted of children attending to Paediatrics Emergency of Porto (UPP). Individuals accompanying those children were approached, in the waiting room in order to answer the designed questionnaire about the patient. Individuals who refused to answer, entered directly to the emergency room or did not complete the interview were excluded. Results: 151 patients entered the emergency room during the shifts, 126 were approached, 25 were immediately non available, 6 refused answering and 8 were incomplete – the response rate was 74%. The age mean was 4 years old, with 46% being under 2. Within children who visited UPP, 37% have been referred: 63% from health centres, 26% from other hospitals and the remaining from private care institutions; 25% of the patients went only to health centres for outpatient care; 56% attended UPP from 2 to 5 times in the last 14 months. Conclusions: Implementing an EPR accessible in health centres and the UPP is relevant for the population studied, especially those under 2 years of age.

1 INTRODUCTION

As information and communication technologies have advanced, interest in mobile health care systems has grown. (Yoo, Kim, Park, Choi, & Chun, 2003).

It happens that a citizen often resorts to more than one medical institution, leading to repeated medical exams and anamnesis. Thus, patient's data is spread over the places where they have received clinical services (Katehakis, Sfakianakis, Tsiknakis, & Orphanoudakis, 2001), such as hospitals, private clinics, pharmacies, etc. (Lambrinoudakis & Gritzalis, 2000). On the other hand, increasing evidence suggests that error in medicine is frequent and may result in substantial harm. (Bates et al., 2001)

To overcome this problem, we need solutions that integrate data (Yoo et al., 2003). Changes have been made, and nowadays, the physical location of a patient record can be replaced by a virtual one (van Bemmel, van Ginneken, Stam, & van Mulligen, 1998), by linking databases from different health institutions attended. When used, computer-based decision supports significantly improved decision quality (Sintchenko, Coiera, Iredell, & Gilbert, 2004) and helped to reduce the frequency and consequences of errors in medical care (Bates et al., 2001 ; Koppel et al., 2005).

These procedures and attempts are expected to allow coordination of information and correction of gaps in communication (considered to be vital (Branger, van't Hooft, & van der Wouden, 1995)) in cases that patients are receiving shared care, which involves more than one physician (Branger, van Hooft, Duisterhout, & van der Lei, 1994) (the number of those cases has increased along time, thanks to the rising of interdisciplinary practice and tests (Hildebrand et al., 2006)). EPRs can also prove themselves important for improvement of quality and reduction of the health's care cost (Kahn, 1997) and improvement of the safety and efficiency of clinical care (Shapiro et al., 2006). EPRs are also believed to be crucial for the creation of large databases of de-identified aggregated data for research (Overhage et al., 2002). However, some studies demonstrated that the application of these new systems may bring some problems, namely the difficulty felt by some emergency physicians to remember their passwords and the time required to search for the information, stressed as significant barriers to access clinical information online (Schneider, 2001).

As far as children are concerned, they are one of the major groups (along with elderly people) who usually request a large amount of health services (INE, 2007) and are involved in emergency situations probably due to the unpredictability of their behaviour. Therefore, the exchange of electronic information regarding a patient who may be unconscious or unaware of his pathological previous condition (cardiac malformation, for example) could prove to be vital in the urgent care scenario.

For that matter, assessing the number of institutions visited by children as well as the proportion that goes to a second health care centre is an issue of major importance as it may allow, in the future, concluding about the number of people that would benefit of such a method.

UPP (Urgência Pediátrica do Porto - Paediatric Unit of Porto) is a centralized healthcare institution for all emergency paediatric events in Porto district.

This study aims to establish the need for the implementation of a regional Electronic Patient Record (EPR) in the Porto area, to support the mobility and the characteristics of the population that attends UPP.

The paper is organized in several sections: a first section of introduction and objective; secondly a participants and methods section in which its described the study, the applied questionnaire and the population; thirdly a results section which presents the results concerning mobility for hospitals and mobility for other institutions; fourthly a discussion section, presenting the authors interpretation of results and study's limitations; next a conclusion section highlighting the main study conclusions; followed by acknowledgments and references.

2 PARTICIPANTS AND METHODS

2.1 Study Design and Population

This is a cross-sectional survey with a consecutive sampling method. Data collection was done in a single moment by interviews to people accompanying children attending the UPP, about events occurred since January of 2006 until the date of interview.

The study population consisted of children attending to Paediatric Emergency of Porto (children under the age of 14).

2.2 Data Collection Methods

The questionnaire was applied to all individuals at the waiting room of UPP, except the ones who refused to answer or those who were immediately non available (entered directly to be attended). Some of the questionnaires were incomplete because, in the mean time, the interviewee was called to receive medical care.

The questionnaire was applied on the 8th, 11th and 12th March (a Sunday and two working days), with two shifts a day, between 8 to 10 a.m. and 8 to 10 p.m. Three interviewers in the waiting room of UPP, approached people who were accompanying children and felt capable of answer our questionnaire. The first stage of the interview was to explain the purposes of the study and ask a consent declaration to be signed (elaborated by the Hospital Ethics Committee). The next stage was to proceed with the questionnaire. The interviews lasted about 5 minutes.

2.3 Questionnaire

Data collected from the interviews included questions about the person who was answering (profession, kinship and years of school) and about the child they were accompanying (age, address,

gender). The remaining questions were targeted to the purpose of the study - type and number of health institutions visited since January 2006 (hospitals, health centres, private clinics) as well as physicians' specialities.

The variables obtained from the questionnaire are age, gender, address, number of each type of healthcare institution attended (hospitals, health centres and private clinics) and also a description of the visited ones. Other variables obtained about the interviewee (the person accompanying the child) are: profession, kinship and years of school.

Age was categorized in age groups. The variables about the number of different healthcare institutions attended were obtained by questioning how many different health care institutions were visited, however a list was provided, being some enumerated to help the individuals remember them. Different institutions not existent in the list were also added.

2.4 Statistical Analysis

The data was submitted to statistic treatment with SPSS 14.0 for windows. Socio-demographic data (age, place of living, employment, qualifications) was presented in frequency distribution; other data, concerning the aim of the study (mobility, referrals, and type of institutions) was mainly submitted to cross frequency tables and absolute frequency distributions.

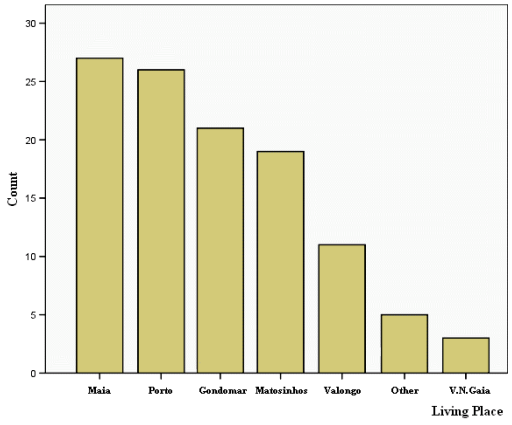


Figure 1: Living places distribution (n=112).

3 RESULTS

3.1 Sample Description

Data was collected in 4 hours shifts, for 3 days, on March 2007. A hundred and fifty one people entered the UPP and, of those, 25 (17%) went directly to the doctor's room. Only 126 (83%) stayed in the waiting room where the interviews took place: 6 (4%) refused, 120 (79%) accepted to answer it. Some of them were called to walk into the doctor's room while answering, so that we obtained 8 (5%) incomplete questionnaires and only 112 (74%) were complete (considered the valid ones). The response rate was 74% (112 out of 151).

The sample was composed by 48 (43%) female and 64 (57%) male children. Fifty one (46%) individuals were under or 2 years old and 61 (54%) were over 2 years old. The mean age was 4 years old. These individuals were living mainly in the Porto region (see Figure 1).

3.2 Mobility for Hospitals

In the study, the mean number of times children attended UPP is 1,74 (σ = 0,61), 39 (35%) children were visiting UPP for the first time since January 2006, 63 (56%) visited it 2, 3, 4 or 5 times and the other 10 (9%) visited it more than 5 times.

Of the individuals who visited UPP, 41 (37%) were referred from other institutions: 27 (63%) from Health Centers, 11 (26%) from Hospitals and 5 (12%) from Private Clinics. The mean of the number of references is 0,49 (σ = 0,76).

Crossing the number of times children attended UPP with the number of references, it's seen that the individuals who visited it more times are the ones who got a higher number of references. However the majority of individuals were not even referred (63%) to other institutions.

Twenty-seven (24%) individuals attended the emergency service of others hospitals and 85 (76%) didn't. Of those who did, 6 (22%) had been at Hospital de Valongo, 3 (11%) at Centro Hospitalar de Vila Nova de Gaia, 3 (11%) at Hospital Geral de Santo António, 1 (4%) at Hospital Pedro Hispano, 1 (4%) at Hospital Maria Pia and 13 (48%) at others.

Thirty-two (29%) of the children needed hospitalization and 80 (71%) didn't. Of the ones who needed it, only 11 (34%) were referred for it and the other 21 (66%) didn't.

3.3 Mobility in Other Health Institutions

Crossing the attendance of the health institutions' data analyzed in three types of institutions (Private Clinics, Health Centers and Outpatient Department) there are 12 (11%) people who attended the three health institutions and only 3 (3%) of these of people did not attend any of those, 28 (25%) of the individuals attended only Health Centers, 14 (12%) only attended Private Clinics, 26 (23%) private clinics and health centres, 17 (15%) health centres and hospitals (see Table 1).

There were 83 (74%) individuals that attended Health Centres within their residence areas. 56 (50%) had visited Private Clinics, mostly Private Consultants for paediatrics; other private clinics include: general clinics, otorhinolaryngology, ophthalmology, orthopaedics, psychology and dermatology were also attended by those children (see Table 2).

When dealing with the outpatient department, there is mobility among various institutions as may be seen in the Table 3.

The distribution of the number of visits to UPP crossed with the age of the children is described in Table 4.

4 DISCUSSION

The interviewees, as expected, were mainly from the Porto region.

Considering that 97% of patients that attended UPP had already gone to at least one medical consultation, it is valid to assume that these individuals would, somehow, benefit from a system that integrates data. Although this value seems to be high, it must be kept in mind that it probably accounts for mandatory vaccination consultations or seeking a second consult concerning the same issue. Comparing the attendance to health centres only (25%) to the attendance to private clinics only (15%), we realized that approximately double the individuals chose health centres over private clinics.

Our results also showed that, within patients who had been referred, a high percentage was referred from health centres (63%). Health centres are also, apart from hospitals, the type of institutions that people most visit in their residence areas (74%). It would then be of special interest to implement an EPR in health centres. Implementing an integrated EPR would allow a rapid and less error-prone information exchange especially in an emergency situation when previous clinical information is of the up-most importance.

Table 1: Cross table of the attendance of Private Clinics, Health Centres and Outpatient Department (n=112).

Attended a Private Clinic	Attended a Health Centre: Yes	Attended a Health Centre: No	Attended Outpatient Department (hospitals)
Yes	12 (11%)	4 (4%)	Yes
	26 (23%)	14 (12%)	No
No	17 (15%)	8 (7%)	Yes
	28 (25%)	3 (3%)	No

Table 2: Table of the number of visits to Health Centers and Private Clinics with its specialities.

Institution	**Cases (n)**
Health Centre	83
Private Clinics	56
Paediatrics	53
General Clinics	4
Otorhinolaryngology	2
Ophthalmology	2
Orthopaedics	2
Psychology	1
Dermatology	1

Table 3: Table of the different institutions attended for outpatient department.

Institution	**Cases (n)**
H. São João	14
H. Pedro Hispano	10
H.G. Santo António	9
H. Maria Pia	4
Others	5
Missing	1

Table 4: Distribution (number of cases and percentage) of the number of visits to UPP by age (n=112).

	Number of visits to UPP			
Age	1st Visit	2-5	>5	Total
≤2	14 (13%)	36 (32%)	1 (1%)	51 (46%)
>2	25 (22%)	27 (24%)	9 (8%)	61 (54%)
Total	39 (35%)	63 (56%)	10 (9%)	112 (100%)

The age factor seems to have some influence in the number of visits to UPP, since individuals until the age of 2 have the higher frequency of 2-5 visits

(32% against 24% of the ones aged over 2). This supports the idea that these individuals should be considered as a target population for the implementation of electronic integrated databases.

Regarding the different institutions chosen by individuals for outpatient care, several are attended; Hospital Pedro Hispano, Hospital Geral Santo António, are the most visited.

When attending private clinics, the most requested speciality is paediatrics, as it was expected.

In a previous similar study done by students of the Faculty of Medicine of University of Porto on elderly population, it was shown that there should be an exchange of information between health institutions in the Porto region. Our study emphasizes this notion because children in this area visit a great number of institution but choose to one emergency care facility (UPP) only; as such, the main conclusion of the study is that integrated EPR between hospitals and health centres would most benefit of children population in Porto region.

4.1 Limitations

The UPP's waiting room, where interviews took place, was itself a limitation to the receptivity of the interviewees due to emotional distress associated with an emergency situation impairing their ability to answer accurately.

The current organization of Porto's paediatric emergency referrals (a centralized healthcare facility for all events), may difficult the generalization to other cities.

5 CONCLUSIONS

Implementing an Electronic Patient Records accessible in health centres and the UPP is relevant for the population studied, especially those under 2 years of age.

ACKNOWLEDGEMENTS

The authors wish to thank Prof. Dr. Almeida Santos (director of UPP) for all the collaboration in this study, and Prof. Dr. Altamiro Costa Pereira for reviewing the article.

This study was supported by POCTI/0753/2004 - Unit I&D CINTESIS nº 753/2004 from Fundação para a Ciência e a Tecnologia.

REFERENCES

Bates, D., Cohen, M., Leape, L., Overhage, J., Shabot, M., & Sheridan, T. (2001). Reducing the frequency of errors in medicine using information technology. *J Am Med Inform Assoc, 8*(4), 299-308.

Branger, P., van Hooft, A., Duisterhout, J., & van der Lei, J. (1994). *A standardized message for supporting shared care.* Paper presented at the Annu Symp Comput Appl Med Care.

Branger, P., van't Hooft, A., & van der Wouden, H. (1995). *Coordinating shared care using electronic data interchange.* Paper presented at the Medinfo.

Hildebrand, C., Pharow, P., Engelbrecht, R., Blobel, B., Savastano, M., & Hovsto, A. (2006). BioHealth--the need for security and identity management standards in eHealth. *Stud Health Technol Inform, 121*, 327-336.

INE. (2007). *Homepage of Instituto Nacional de Estatística*. Retrieved 2007 May 17, from http://www.ine.pt/prodserv/quadros/periodo.asp?pub_cod=260

Kahn, M. (1997). Three perspectives on integrated clinical databases. *Acad Med, 72*(4), 281-286.

Katehakis, D., Sfakianakis, S., Tsiknakis, M., & Orphanoudakis, S. (2001). An infrastructure for Integrated Electronic Health Record services: the role of XML (Extensible Markup Language). *J Med Internet Res, 3*(1), E7.

Koppel, R., Metlay, J., Cohen, A., Abaluck, B., Localio, A., Kimmel, S., et al. (2005). Role of computerized physician order entry systems in facilitating medication errors. *JAMA, 293*(10), 1197-1203.

Lambrinoudakis, C., & Gritzalis, S. (2000). Managing medical and insurance information through a smart-card-based information system. *J Med Syst, 24*(4), 213-234.

Overhage, J. M., Dexter, P. R., Perkins, S. M., Cordell, W. H., McGoff, J., McGrath, R., et al. (2002). A Randomized, Controlled Trial of Clinical Information Shared From Another Institution. *Ann Emerg Me, 39*(1), 14-23.

Schneider, J. (2001). Online personal medical records: are they reliable for acute/critical care? . *Crit Care Med, 29*(8 Suppl), N196-201.

Shapiro, J., Kannry, J., Lipton, M., Goldberg, E., Conocenti, P., Stuard, S., et al. (2006). Approaches to patient health information exchange and their impact on emergency medicine. *Ann Emerg Med, 48*(4), 426-432

Sintchenko, V., Coiera, E., Iredell, J., & Gilbert, G. (2004). Comparative impact of guidelines, clinical data, and decision support on prescribing decisions: an interactive web experiment with simulated cases. *J Am Med Inform Assoc, 11*(1), 71-77.

van Bemmel, J., van Ginneken, A., Stam, B., & van Mulligen, E. (1998). *Virtual Electronic patient records for shared care.* Paper presented at the Medinfo.

Yoo, S., Kim, B., Park, H., Choi, J., & Chun, J. (2003). *Realization of real-time clinical data integration using advanced database technology.* Paper presented at the AMIA Annu Symp.

AUTHOR INDEX

AUTHOR INDEX (CONT.)

Proceedings of HEALTHINF 2008
International Conference on Health Informatics
ISBN: 978-989-8111-16-6