Health Informatics and Knowledge Management 2011

Health Informatics and Knowledge Management 2011

Proceedings of the Fourth Australasian Workshop on
Health Informatics and Knowledge Management
(HIKM 2011), Perth, Australia,
17-20 January 2011

Kerryn Butler-Henderson and Tony Sahama, Eds.

Health Informatics and Knowledge Management 2011. Proceedings of the Fifth Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2011), Perth, Australia, 17-20 January 2011
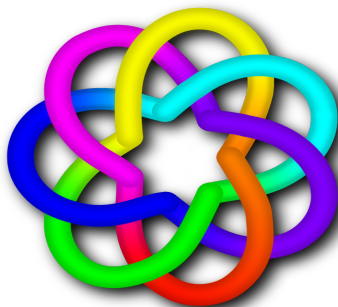
Conferences in Research and Practice in Information Technology, Volume 120.

Editors:

Kerryn Butler-Henderson
Department of Health Information Management
School of Public Health
Curtin University
Perth, WA 6845
Australia
Email: `K.Butler-Henderson@exchange.curtin.edu.au`

Tony Sahama
Faculty of Science and Technology (FaST)
Qeensland University of Technology
Brisbane, Queensland 4001
Australia
Email: `t.sahama@qut.edu.au`

Series Editors:
Vladimir Estivill-Castro, Griffith University, Queensland
Simeon J. Simoff, University of Western Sydney, NSW
Email:`crpit@scm.uws.edu.au`

Table of Contents

Contributed Papers

Preface

We are pleased to present the papers from the Australasian Health Informatics and Knowledge Management (HIKM) conference stream held on 20 January 2011 in Perth as a session of the Australasian Computer Science Week (ASCW) 2011. Formerly HIKM was named Health Data and Knowledge Management, however the inclusion of the health informatics term is timely given the current health reform. The submissions to HIKM 2011 demonstrated that Australasian researchers lead with many research and development innovations coming to fruition. Some of these innovations can be seen here, and we believe further recognition will accomplish by continuation to HIKM in the future.

The HIKM conference is a review of health informatics related research, development and education opportunities. The conference papers were written to communicate with other researchers and share research findings, capturing each and every aspect of the health informatics field. They are namely: conceptual models and architectures, privacy and quality of health data, health workflow management patient journey analysis, health information retrieval, analysis and visualisation, data integration/linking, systems for integrated or coordinated care, electronic health records (EHRs) and personally controlled electronic health records (PCEHRs), health data ontologies, and standardisation in health data and clinical applications.

There was a high standard of submissions for HIKM, with nine of the 21 submissions selected following a detailed peer review by three independent expert reviewers. We hope the unsuccessful authors will seek other opportunities to present their work as this body of knowledge needs to be shared with the Australasian research community. The nine papers selected are from three different countries, presenting a variety of topics from the theoretical to the clinical. We congratulate those who presented at HIKM 2011.

Kerryn Butler-Henderson
Curtin University

Tony Sahama
Queensland University of Technology

HIKM 2011 Programme Chairs
January 2011

Acknowledgement

Programme Committee

Chairs

Kerryn Butler-Henderson, Curtin university, Australia
Tony Sahama, Queensland University of Technology, Australia

Members

Vicki Bennett, Australian Institute of Health and Welfare; Health Information Management Association of Australia, Australia
David Hansen, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia
Travis Ingram, Department of Health WA, Australia
Mitchell Jesson, Department of Health WA, Australia
Steve Kisely, university of Queensland, Australia
Keith Liu, University of Western Australia, Australia
Anthony Maeder, University of Western Sydney, Australia
Timothy Nelson, Curtin University, Australia
Louise Schaper, Health Informatics Society of Australia, Australia
Andrew Stranieri, University of Ballarat, Australia
Jim Warren, University of Auckland, New Zealand
Dennis Wollersheim, La Trobe University, Australia
Ping Yu, University of Wollongong

Organising Committee

Chair

Assoc. Prof. Mihai Lazarescu

Co-Chair

Assoc. Prof. Ling Li

Finance

Mary Simpson
Mary Mulligan

Catering and Booklet

Mary Mulligan
Dr. Patrick Puersum
Assoc. Prof. Mihai Lazarescu

Sponsorship and Web

Dr. Patrick Peursum
Dr. Aneesh Krishna

Registration

Mary Mulligan
Dr. Patrick Puersum

DVD and Signage

Dr. Patrick Peursum
Mary Mulligan

Venue

Dr. Mike Robey

Conference Bag

Dr. Sieteng Soh

Welcome from the Organising Committee

On behalf of the Australasian Computer Science Week 2011 (ACSW2011) Organising Committee, we welcome you to this year's event hosted by Curtin University. Curtin University's vision is to be an international leader shaping the future through its graduates and world class research. As Western Australia's largest university, Curtin is leading the state in producing high quality ICT graduates. At Curtin Computing, we offer both world class courses and research. Our Computing courses cover three key areas in IT (Computer Science, Software Engineering and Information Technology), are based on the curricula recommendations of IEEE Computer Society and ACM, the largest IT professional associations in the world, and are accredited by the Australian Computer Society. Curtin Computing hosts a top level research institute (IMPCA) and offers world class facilities for large scale surveillance and pattern recognition.

We welcome delegates from over 18 countries, including Australia, New Zealand, USA, U.K., Italy, Japan, China, Canada, Germany, Spain, Pakistan, Austria, Ireland, South Africa, Taiwan and Thailand. We hope you will enjoy the experience of the ACSW 2011 event and get a chance to explore our wonderful city of Perth. Perth City Centre is located on the north bank of the Swan River and offers many fun activities and a wealth of shopping opportunities. For panoramic views of Perth and the river, one can visit Kings Park or enjoy a relaxing picnic in one of the many recreational areas of the park.

The Curtin University campus, the venue for ACSW2011, is located just under 10km from the Perth City Centre and is serviced by several Transperth bus routes that travel directly between Perth and Curtin University Bus Station, as well as several other routes connecting to nearby train services.

ACSW2011 consists of the following conferences:

- Australasian Computer Science Conference (ACSC) (Chaired by Mark Reynolds)
- Australasian Computing Education Conference (ACE) (Chaired by John Hamer and Michael de Raadt)
- Australasian Database Conference (ADC) (Chaired by Heng Tao Shen and Athman Bouguettaya)
- Australasian Information Security Conference (AISC) (Chaired by Colin Boyd and Josef Piepryzk)
- Australasian User Interface Conference (AUIC) (Chaired by Christof Lutteroth)
- Australasian Symposium on Parallel and Distributed Computing (AusPDC) (Chaired by Jinjun Chen and Rajiv Ranjan)
- Australasian Workshop on Health Informatics and Knowledge Management (HIKM) (Chaired by Kerryn Butler-Henderson and Tony Sahama)
- Computing: The Australasian Theory Symposium (CATS) (Chaired by Taso Viglas and Alex Potanin)
- Australasian Computing Doctoral Consortium (ACDC) (Chaired by Rachel Cardell-Oliver and Falk Scholer).

The nature of ACSW requires the co-operation of numerous people. We would like to thank all those who have worked to ensure the success of ACSW2011 including the Organising Committee, the Conference Chairs and Programme Committees, our sponsors, the keynote speakers and the delegates. Many thanks go to Alex Potanin for his extensive advice and assistance and Wayne Kelly (ACSW2010 chair) who provided us with a wealth of information on the running of the conference. ACSW2010 was a wonderful event and we hope we will live up to the expectations this year.

Assoc. Prof. Mihai Lazarescu and Assoc. Prof. Ling Li
Department of Computing, Curtin University

ACSW2011 Co-Chairs
January, 2011

CORE - Computing Research & Education

CORE welcomes all delegates to ACSW2011 in Perth. CORE, the peak body representing academic computer science in Australia and New Zealand, is responsible for the annual ACSW series of meetings, which are a unique opportunity for our community to network and to discuss research and topics of mutual interest. The original component conferences ACSC, ADC, and CATS, which formed the basis of ACSWin the mid 1990s now share this week with six other events - ACE, AISC, AUIC, AusPDC, HIKM, ACDC, which build on the diversity of the Australasian computing community.

In 2011, we have again chosen to feature a small number of plenary speakers from across the discipline: Heng To Shen, Gene Tsudik, ans Dexter Kozen. I thank them for their contributions to ACSW2011. I also thank the keynote speakers invited to some of the individual conferences. The efforts of the conference chairs and their program committees have led to strong programs in all the conferences again, thanks. And thanks are particularly due to Mihai Lazarescu and his colleagues for organising what promises to be a strong event.

In Australia, 2009 saw, for the first time in some years, an increase in the number of students choosing to study IT, and a welcome if small number of new academic appointments. Also welcome is the news that university and research funding is set to rise from 2011-12. However, it continues to be the case that per-place funding for computer science students has fallen relative to that of other physical and mathematical sciences, and, while bodies such as the Australian Council of Deans of ICT seek ways to increase student interest in the area, more is needed to ensure the growth of our discipline.

During 2010, CORE continued to negotiate with the ARC on journal and conference rankings. A key aim is now to maintain the rankings, which are widely used overseas as well as in Australia. Management of the rankings is a challenging process that needs to balance competing special interests as well as addressing the interests of the community as a whole.

COREs existence is due to the support of the member departments in Australia and New Zealand, and I thank them for their ongoing contributions, in commitment and in financial support. Finally, I am grateful to all those who gave their time to CORE in 2010; in particular, I thank Alex Potanin, Jenny Edwards, Alan Fekete, Aditya Ghose, Leon Sterling, and the members of the executive and of the curriculum and ranking committees.

Tom Gedeon

President, CORE
January, 2011

ACSW Conferences and the
Australian Computer Science Communications

The Australasian Computer Science Week of conferences has been running in some form continuously since 1978. This makes it one of the longest running conferences in computer science. The proceedings of the week have been published as the *Australian Computer Science Communications* since 1979 (with the 1978 proceedings often referred to as *Volume 0*). Thus the sequence number of the Australasian Computer Science Conference is always one greater than the volume of the Communications. Below is a list of the conferences, their locations and hosts.

2012. Volume 34. Host and Venue - RMIT University, Melbourne, VIC.

2011. **Volume 33. Host and Venue - Curtin University of Technology, Perth, WA**.

2010. Volume 32. Host and Venue - Queensland University of Technology, Brisbane, QLD.
2009. Volume 31. Host and Venue - Victoria University, Wellington, New Zealand.
2008. Volume 30. Host and Venue - University of Wollongong, NSW.
2007. Volume 29. Host and Venue - University of Ballarat, VIC. First running of HDKM.
2006. Volume 28. Host and Venue - University of Tasmania, TAS.
2005. Volume 27. Host - University of Newcastle, NSW. APBC held separately from 2005.
2004. Volume 26. Host and Venue - University of Otago, Dunedin, New Zealand. First running of APCCM.
2003. Volume 25. Hosts - Flinders University, University of Adelaide and University of South Australia. Venue - Adelaide Convention Centre, Adelaide, SA. First running of APBC. Incorporation of ACE. ACSAC held separately from 2003.
2002. Volume 24. Host and Venue - Monash University, Melbourne, VIC.
2001. Volume 23. Hosts - Bond University and Griffith University (Gold Coast). Venue - Gold Coast, QLD.
2000. Volume 22. Hosts - Australian National University and University of Canberra. Venue - ANU, Canberra, ACT. First running of AUIC.
1999. Volume 21. Host and Venue - University of Auckland, New Zealand.
1998. Volume 20. Hosts - University of Western Australia, Murdoch University, Edith Cowan University and Curtin University. Venue - Perth, WA.
1997. Volume 19. Hosts - Macquarie University and University of Technology, Sydney. Venue - Sydney, NSW. ADC held with DASFAA (rather than ACSW) in 1997.
1996. Volume 18. Host - University of Melbourne and RMIT University. Venue - Melbourne, Australia. CATS joins ACSW.
1995. Volume 17. Hosts - Flinders University, University of Adelaide and University of South Australia. Venue - Glenelg, SA.
1994. Volume 16. Host and Venue - University of Canterbury, Christchurch, New Zealand. CATS run for the first time separately in Sydney.
1993. Volume 15. Hosts - Griffith University and Queensland University of Technology. Venue - Nathan, QLD.
1992. Volume 14. Host and Venue - University of Tasmania, TAS. (ADC held separately at La Trobe University).
1991. Volume 13. Host and Venue - University of New South Wales, NSW.
1990. Volume 12. Host and Venue - Monash University, Melbourne, VIC. Joined by Database and Information Systems Conference which in 1992 became ADC (which stayed with ACSW) and ACIS (which now operates independently).
1989. Volume 11. Host and Venue - University of Wollongong, NSW.
1988. Volume 10. Host and Venue - University of Queensland, QLD.
1987. Volume 9. Host and Venue - Deakin University, VIC.
1986. Volume 8. Host and Venue - Australian National University, Canberra, ACT.
1985. Volume 7. Hosts - University of Melbourne and Monash University. Venue - Melbourne, VIC.
1984. Volume 6. Host and Venue - University of Adelaide, SA.
1983. Volume 5. Host and Venue - University of Sydney, NSW.
1982. Volume 4. Host and Venue - University of Western Australia, WA.
1981. Volume 3. Host and Venue - University of Queensland, QLD.
1980. Volume 2. Host and Venue - Australian National University, Canberra, ACT.
1979. Volume 1. Host and Venue - University of Tasmania, TAS.
1978. Volume 0. Host and Venue - University of New South Wales, NSW.

Conference Acronyms

ACDC	Australasian Computing Doctoral Consortium
ACE	Australasian Computer Education Conference
ACSC	Australasian Computer Science Conference
ACSW	Australasian Computer Science Week
ADC	Australasian Database Conference
AISC	Australasian Information Security Conference
AUIC	Australasian User Interface Conference
APCCM	Asia-Pacific Conference on Conceptual Modelling
AusPDC	Australasian Symposium on Parallel and Distributed Computing (replaces AusGrid)
CATS	Computing: Australasian Theory Symposium
HIKM	Australasian Workshop on Health Informatics and Knowledge Management

Note that various name changes have occurred, which have been indicated in the Conference Acronyms sections in respective CRPIT volumes.

ACSW and HIKM 2011 Sponsors

We wish to thank the following sponsors for their contribution towards this conference.

CORE - Computing Research and Education,
www.core.edu.au

Curtin University of Technology,
www.curtin.edu.au

Perth Convention Bureau,
www.pcb.com.au

Australian Computer Society,
www.acs.org.au

CONTRIBUTED PAPERS

Decision Support Based Needs Assessment for Cancer Patients

Andrew Stranieri[1], Siddhivinayak Kulkarni[1], Alyx Macfadyen[2], Anthony Love[3], Stephen Vaughan[4]

[1]Centre for Informatics and Applied Optimisation, University of Ballarat, Australia
[2]LaTrobe University, Bendigo, Australia
[3]School of Behavioral and Social Sciences, Univesity of Ballarat , Australia
[4]Grampians Integrated Cancer Services, Ballarat, Australia

a.stranieri@ballarat.edu.au

Abstract

Regular assessment of wellness or quality of life for patients throughout a cancer journey is important so as to identify aspects of life that could lead to distress and impede recovery or acceptance. The emerging trends in assessment are to deploy validated, quality of life instruments on touchscreen computers in medical waiting rooms. However, these add to workload of health care professionals and can be impersonal for patients to use. In this article, an alternate approach is presented that involves a decision support system with natural dialogue that elicits the patient's specific context in a far finer grained manner than is possible with questionnaire based instruments. The system includes a model of heuristics that health care professionals in a locality use to make inferences regarding a patient's quality of life and avenues for referral.

Keywords: Needs assessment, Decision support, Supportive care, Cancer

1 Introduction

Numerous studies and reports have identified that many cancer patients have concerns with transport, finances, relationships, spirituality, pain, home, personal care and other facets of life over and above medical issues. Further, concerns impact on the morale of the patient and threaten to compromise positive attitudes to treatment or acceptance of the disease. (Gerber 2001)

Currently, the medical profession does not systematically include assessments of a patient's overall well-being and concerns into patient care. Specialist health care professionals serving cancer patients are typically faced with enormous time constraints.

Workflows and the realities of medical specialisations are designed to optimise human resources and often make it difficult for professionals to get to know patients sufficiently well to make an assessment of concerns outside the sphere of their medical speciality. The approach currently emerging in response to this problem involves the development of short questionnaire based instruments, rigorously validated and presented to patients while they wait for professionals or treatment. The first of these instruments to be used in Australia was developed by Foot (1995) and was to used to specifically evaluate the needs of cancer patients. The Cancer Needs Questionnaire (CNQ) identified three levels of need (low, moderate, high) and two levels of no-need (no need, need satisfied). Since, then many other instruments, surveyed below, that have been validated to various degrees of rigor, have emerged.

The instruments are typically paper based and provided to patients to complete while waiting for a health care professional. However, processing the questionnaire data adds to the workload of health care professionals so has proven to be less than popular. The response to this has been to develop computer based versions with easy to use touch screens installed in waiting rooms for patients to use prior to a consultation.

The central claim of this paper is that the use of validated instruments and technological devices installed in waiting rooms, while possibly effective, does not put cancer patients at the centre of their treatment. Prompting patients who have a life threatening illness to indicate intimate personal details such as their level of distress, the quality of their sex life, or their reaction to being increasingly burdensome on their carer to a computer screen mounted on the wall of a waiting room does not obviously empower patients.

In this article, we advance an alternative approach where heuristics that health care professionals use to make a judgement of a patient's well being are modelled and embedded in a web based decision support system that patient's can access from their homes, waiting rooms or mobile devices. The system incorporates question and answer sequences that are designed to be as natural and sensitive to a patient's specific context as possible.

The system is designed to be very readily extensible so health care professionals (or patients) who think of a need currently not catered for, can suggest questions that

patients might benefit from being asked, and have new questions-answer sequences entered. The approach can be conceived of as a Wikipedia style approach where the system evolves with user input and stands in contrast to the Britannica approach, more inflexible yet more rigorously validated.

A conventional decision support technology based on rule based representation is not sufficiently flexible for the extensibility required because any suggestion for a change in content would need to be executed by specialist knowledge engineers. Instead a relatively simple approach described below that permits the flexibility required, has been developed.

Currently, the system has been implemented and populated with rudimentary question-answer sequences. A system infers a snapshot of the state of wellness of a patient using basic criteria that can readily be extended. The system also infers recommendations for actions the patient or health care professional can take. The patient is expected to use the web based system on the internet in the comfort of their home and ultimately grant access rights to their health care professionals or take a print out of the system inference to their next visit. In this way, the report becomes a useful addition to the health care professional's tasks without adding to their workload and importantly, keeps the patient at the centre of their treatment.

This article reports on progress to date with this approach. The next phase of this project involves a trial to gauge the extent to which health care professionals and patients are likely to benefit from the systems inferences and contribute to its expansion. In the next section of this paper a review of the literature that identifies unmet needs as an issue and quality of life and other instruments that have been advanced in response to the need, is presented. Following that, the design of the system and pilot is described.

2   Wellness, needs and quality of life

Gerber (2001) identified that cancer hospital inpatients had diverse needs which are best understood in order to promote an improvement in the quality of care. Wen and Gustafson (2004) also emphasise the importance of an assessment of patient needs, satisfaction and quality of life. They reviewed the assessment instruments of needs experienced by cancer patients and their families and conclude that a careful assessment of patients' needs is central to the whole process of providing care.

According to Siegal (1994), cancer patients can encounter financial problems as income often diminishes while medical expenses increase. Further, (Houts et al 1984) discovered that as a patient's functional capacity diminishes, out-of-pocket expenses rise due to an increase in needs for necessities such as transportation, child care, home care services, extra or specific foods, clothing, and lost income and wages. Berkman (1993) also found that out-of-pocket expenses coupled with gaps in medical insurance coverage cause long-term financial difficulty and jeopardize patients' ability to obtain or keep health insurance coverage leading to spiralling problems. Mor et al (1987) found that 41 percent of households studied reported a 50% increase in expenses and 16 percent of households studied were unable to meet their monthly bills.

Bryan et al (1991) reported that cancer patients face a variety of problems related to transportation including availability of, and access to public transport, difficulties associated with driving or having others drive and trvel For patients in regional areas, travel can require accommodation, which involves logistic challenges. Among elderly cancer patients, Goodwin et al (2003) found that those with reduced access to transportation were four times less likely to receive radiation therapy following surgical treatment.

Financial and transport concerns are tangible and relatively easily recognised by patients themselves. Other issues may not be so readily apparent but no less concerning. Surbone and Baider (2010) discuss the spiritual dimension involving cancer patients. Rainbird et al (2005) identified seven categories of needs for the cancer patient: Medical communication and information, Psychological/emotional, Daily living, Financial, Dealing with symptoms, Spiritual and Social.

A number of instruments have been developed in recent years to assess the diversity of needs. Foot et al. (1995) developed the first Australian Survey to evaluate the needs of the cancer patients known as the Cancer Needs Questionnaire (CNQ). Another commonly used questionnaire is the Supportive Care Needs Survey advanced by Bonevski (2000). This survey includes issues related to sexuality. The questionnaire developed by Osse (2005) is tailored for palliative care: the Problems and Needs in Palliative Care questionnaire (PNPC).

The Cancer Rehabilitation Evaluation System (CARES) advanced by Schag (1990) measures need items from six domains, first on a five-point scale to indicate whether the need applies and then whether additional help is required. Tamburini et al (2000) developed a Needs Evaluation Questionnaire (NEQ) to evaluate the informative, psychological, social and practical needs of hospitalised cancer patients that would help the medical staff as well as the cancer patients. The NEQ is a standardized questionnaire, consisting of 23 items that has been empirically validated. It assesses the following needs: information concerning diagnosis or prognosis, information concerning treatment, communication, home assistance, financial and other support.

The most pressing observation to make from the plethora of instruments that have been developed is that they vary based on the needs covered, the intended end user and the patient context. This strongly suggests that a single instrument is unlikely to be well suited to every context involving the practice of health care with cancer patients. For instance, health care professionals in remote areas, may require that transport segment of instruments be more sophisticated than in city areas. In localities characterised with patients from high populations of minority groups and practitioners from majority cultures, psycho-social assessments focused on cultural issues may be pertinent in a way that they are not in homogeneous localities.

The development of a validated instrument for each locality is impractical and too rigid in that preferences of

health care professionals and patients in a community change over time. An alternative to developing a validated instrument customised for each locality, or a universal one specific to no single locality, is to develop a decision support system designed to evolve locally to suit the characteristics of each community over time. This does not mean that empirical validation of the decision support system is not important but only that validation studies need to be postponed until the decision support system's knowledge base has been populated to a point where there is widespread albeit informal agreement that the system attempts to tap into the needs that health care professionals and patients in a locality consider important. In the next section, the system developed for this purpose is described. The system currently has been seeded with initial question and answer sequences and some referral heuristics. The next phase of the study will involve refinement of the prototype in a test locality prior to wider deployment.

3 NADSS Needs Assessment Decision Support System

The needs assessment approach adopted in this study involves the development of a decision support system by capturing the heuristics health care professionals deploy when engaging with patients to make an assessment. Ultimately, health care professionals engage patients in a dialogue and apply heuristics that have evolved with experience to determine appropriate questions to ask given a patient's previous answers and finally, advance an assessment regarding the patient's needs.

Koers et al (1989) had found that the sequence of questions an expert asks is more important than the logical structure of knowledge in a domain for the user acceptability of a system. User acceptance is important for any software system but is particularly critical in the field of needs assessment for cancer patients because patients will be expected to take themselves through a consultation, perhaps many times, during their cancer journey. However, acceptance is made difficult because patients may tire easily; they may have heightened sensitivity to quality of life prompts and may have limited computer experience or skills.

Kawamoto et al (2005) identify critical success factors in the successful deployment of clinical decision support systems that include: useability, integration into clinician's existing workflow, provision of a recommendation, integration with other systems and provision of results to patients and clinicians. The current approach is designed to accommodate each of these factors with the exception of the integration with other systems. The transfer of data between health care professionals systems and the needs system, and vice-versa requires that flexibility in databases that most hospital systems currently do not have. In the future, a ubiquitous electronic health record will facilitate this integration.

The current approach involved capturing the question and answer sequences that health care professionals were likely to use in discursive engages with patients leading to heuristic assessments of their needs. Three of the investigators who collectively had many years experience in modelling heuristics and caring for patients derived

question and answer sequences that seemed plausible. The first question prompts the patient to select a need category labelled *'Talking Points'* as illustrated in Figure 1. Categories defined to date were: Financial, Physical, Home/Family, Spiritual, Transport and Social.

The first prompt in the Physical category is:
What would you like to talk about?
- Lets talk about pain
- Lets talk about personal care

Figure 1 Talking point menu

If <u>Lets talk about pain</u> is selected the following prompt appears:
Are you in pain right now?
- I would be if I wasn't managing it
- I am in some pain right now
- No, not really

If <u>I would be if I wasn't managing it</u> was selected then the following prompt is presented:
Which of these applies to you more?
- Its increasingly difficult to manage the pain
- Its currently under control
- Its currently under control but I'm concerned about coping in the future

The consultation continues until an end point is reached where the patient is returned to the *Talking points* menu where another category or a system exit can be selected. On exit, the system performs a needs assessment and generates a report as illustrated in Figure 2.

The question and answer sequence is conceptually equivalent to a decision tree where nodes represent questions and arcs emanating from each node represent alternate answers. This is conceptually equivalent to sets of IF-THEN rules however, the decision support system advanced here varies the traditional approach. Instead of encoding a final assessment in the leaf or conclusion nodes of the tree, the approach advanced here sets variables associated with each arc to values specified by the health care professional and knowledge engineer. These variables represent something about the state of the

patient. For instance, a variable *Anxiety,* initially set to 0, is incremented by 1 when the user responds that he or she would be pain if they were not managing it. The heuristic used here involves the notion that a small level of anxiety is associated with the management of pain.

On conclusion of the consultation, the system compares the variable's values to threshold values linked to suggested actions. For example, a total *Anxiety* score between 3 and 6 triggers the suggestion that the patient should be referred to a social worker. A total anxiety score greater than 6 triggers the suggestion that the patient should be referred to a consultant. The **Snapshot** section of the report produced depicted in Figure 2 derives directly from the variable scores and the **Suggestion** section derives from comparing the scores to thresholds associated with suggested actions.
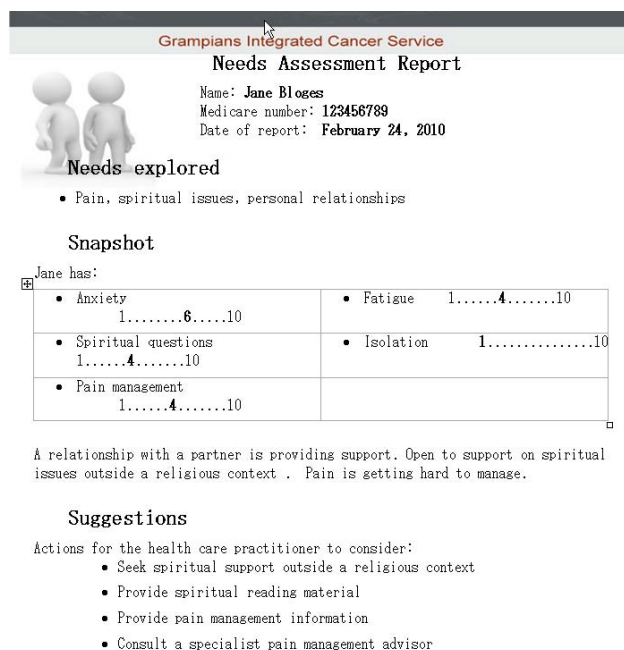

Figure 2 Needs assessment report

This Snapshot and Suggested action mechanism was designed for the following reasons:

- Variables that represent a snapshot of the patient's state were considered useful for health care professionals to quickly gain an overview of a patient's state;
- The suggested actions encode heuristics that experienced professionals use to trigger possible remedial actions. This inference provides health care professionals that enhances and facilitates their role. This is in contrast to questionnaires that identify a need but assume the professional will draw the appropriate inference in inferring a remedial action. This adds cognitive and time load on the professional;
- Encoding heuristics as Snapshot variables and Suggested action heuristics permits the question and answer sequences to be designed to be as natural as possible;

- The approach is readily extensible. Knowledge engineers can associate any number of variables to each answer and assign a numerical value to any variable to represent the extent to which that state is enhanced by the selected answer.

Four modules were deployed within the Patient Needs Assessment System. These were:

- the Expert Knowledge base,
- the Inferencing or Decision Support system,
- the User Interface and;
- the Content Management System.

Expert knowledge guided the content and direction of the questions that patients consider.

The Inferencing System uses two matrices of variables for each category domain. In the first, a score is associated with a variable for each user selected response to a question as presented in Table 1 .

It is increasingly difficult to manage the pain	
Variable	**Score**
Pain	5
Anxiety	2
Fatigue	1
Close others	1

Table 1: Variables and scores on a pain response

The heuristics captured by the investigators as knowledge engineers illustrated in Table 1 involve the belief that a patient finding it increasingly difficult to manage pain will have a little anxiety, a little fatigue and will place an addition strain on others close to them.

Heuristics regarding which professional, if any, the patient should be referred to, is captured as a score on each variable as illustrated in Table 2. The actions in the table depend heavily on the local context. In a region without pain specialists, health care professionals may refer patients to local psychologists so they would simply change the entry in the table.

The User Interface deploys the Bellaswan Content Management System (www.bellaswan.com.au). Some handcrafting was necessary at the higher application interface level to ensure the categories and conversation topics are each presented in a way that fosters trust from both patient and professional. A template thereafter uses a call-back method that populates the lower-level decision tree conversation points. The consistency of content placement is important. If system-generated content is misplaced, inconsistent or overlaps other content, the perception of quality is compromised. Reassurance of

how data that is collected from users must also be clear to patients and professionals from the outset.

Referral inference for the Pain variable	
Score	**Health Care Professional**
7 or more	Recommend a pain specialist
Between 3 and 7	Recommend a GP
Between 1 and 3	Provide pain management information
Less than 1	No action required

Table 2: Referral scores for the variable Pain

The Content Management System (CMS) retrieves, stores, processes and returns data to the interface. As the patient explores the conversation topics and makes selections, a snapshot of their own "Story so far" is displayed on-screen and can be printed or emailed at any time. This acts as a narrative summary of the current consultation.

4   Conclusion

Many instruments have been developed and validated to elicit the physical, financial, transport, psychological, social, sexual, spiritual, pain management and informational needs of cancer patients. To reduce the workload on health care professionals many instruments are being converted to touchscreen, computer implementations installed in waiting rooms. However, these can be uncomfortable for patients to access and do not cater for preferences that professionals or patients have in local communities. An alternative approach is presented that involves the development of a decision support system that presents the patient with a question and answer dialogue that is more natural and can be tailored to suit the context of the community and patient. The decision support system is designed to be very easily extended so that, over time, it evolves to suit local preferences. A system has been designed and populated with an initial knowledge base. Further research will explore the extent to which patients and health care professionals can enhance the knowledge base and derive benefits without adding to the workload of current staff.

5   References

Babcock, B., Datar M., Motwani, R. and O'Callaghan, L. (2003). Maintaining Variance and k-Medians over Data Stream Windows, *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems.* 234-243.

Berkman BJ, Sampson SE. Psychosocial effects of cancer economics on patients and their families. Cancer 1993;72:2846–2849.

Bonevski B, Sanson-Fisher RW, Girgis A, Burton L, Cook BA (2000) Evaluation of an instrument to assess the needs of patients with cancer. Cancer 88:217–225

Bryan JL, Greger HA, Miller ME, et al. An evaluation of the transportation needs of disadvantaged cancer patients. J Psychosoc Oncol 1991;9:23–35.

Bulut, A., Singh, A. (2003). Hierarchical Stream Summarization in Large Networks. *International Conference on Data Engineering.*

Cancer Care, Inc. Evaluating the barriers to pain relief. A Report to the United Hospital Fund 1999

Charikar, M., Chen, K., Farach-Colton, M. (2002). Finding Frequent Items in Data Streams. Lecture Notes in Computer Science. Proceedings of the 29th International Colloquium on Automata, Languages and Programming. (2380) 693-703.

Cossich, T., Schofield, P and McLachlan, S. A., Validation of the cancer needs questionnaire (CNQ) short-form version in an ambulatory cancer setting, Quality of Life Research 13: 1225–1233, 2004.

Degligiannakis, A., Garofalakis, M. and Roussopoulos, N. (2007). Extended Wavelets for Multiple Measures. *ACM Transactions on Database Systems.* 32 (2).

Fong, J., Strauss, M. (2000). An Approximate L$_p$-difference Algorithm for Massive Data Streams. *Lecture Notes in Computer Science. Proceedings of the 17th Annual Symposium on Theoretical Aspects of Computer Science.* (1770) 193-204.

Foot G, Sanson-Fisher R (1995) Measuring the unmet needs of people living with cancer. Cancer Forum 19:131–135

Gaber, M.M. , Zaslavsky, A. and Krishnaswamy, S. (2005). Mining Data Streams: A Review. *ACM SIGMOD Record.* 34(2): 18-26.

Gates MF, Lackey NR, White MR (1995) Needs of hospice and clinic patients with cancer. Cancer Pract 3:226–232

Gerber LH: Cancer Rehabilitation into the future. Cancer 2001 , 92(S):975-9.

Goodwin JS, Hunt WC, Samet JM. Determinants of cancer therapy in elderly patients. Cancer 1993;72:594–601.

Guidry JJ, Aday LA, Zhang D, Winn RJ, Transportation as a barrier to cancer treatment, Cancer Pract. 1997 Nov-Dec;5(6):361-6.

Houts PS, Lipton A, Harvey HA, et al. Nonmedical costs to patients and their families associated with outpatient chemotherapy. Cancer 1984;53:2388–2392.

Indyk, P., Koudas, N., and Muthukrishnan, S. (2000). Identifying Representative Trends in Massive time Series Data Sets Using Sketches. *Proceedings of the 26th International Conference on Very Large Data Bases.* 363- 372.

Kawamoto, K., Houlihan, C, A E., Balas, A., Lobach , D. (2005) Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ Online p1-8. BMJ, doi:10.1136/bmj.38398.500764.8F (published 14 March 2005)

Koers, A. W., Kracht, D., Smith, M., Smits, J. M. and Weusten, M. C. M. 1989. Knowledge Based Systems in Law. Computer/Law Series. Kluwer. The Netherlands.

Lintz K, Moynihan C, Steginga A, Norman A, Eeles R, Huddart R et al (2003) Prostate cancer patients' support and psychological care needs: Survey from a non-surgical oncology clinic. Psychooncology 12:769–783

M. E. McDowell, S. Occhipinti, M. Ferguson, J. Dunn, S. K. Chambers, Predictors of change in unmet supportive care needs in cancer, Psycho-Oncology, Volume 19, Issue 5, pages 508–516, May 2010.

Mor V, Guadagnoli E, Wool M. An examination of the concrete service needs of advanced cancer patients. J Psychosoc Oncol 1987;5:1–17.

O'Hare PA, Malone D, Lusk E. Unmet needs of black patients with cancer posthospitalization: A descriptive study. Oncol Nurs Forum 1993;20:659–664.

Osse B, ., Vernooij-Dassen M, Schade E, Grol R., (2004) Towards a new clinical tool for needs assessment in palliative care of cancer patients: The PNPC instrument. J Pain Symptom Manage 28:329–341

Rainbird, K., Perkins, J., and Sanson-Fisher. (2005). The Needs Assessment for Advanced Cancer Pastients (NA-ACP): A Measure of the Perceived Needs of Patients with Advanced, Incurable Cancer. A Study of Validity, Reliability and Acceptability. *Psycho-Oncology* 14: 297–306

Schag CAC, Heinrich RL (1990) Development of a comprehensive quality of life measurement tool: CARES. Oncology 4:135–138

Siegel K, Raveis VH, Mor V, et al. The relationship of spousal caregiver burden to patient disease and treatment-related conditions. Ann Oncol 1991;2:511–16.

Surbone A, and Baider, L. (2010) The spiritual dimension of cancer care. Critical Reviews in Oncology/Hematology Volume 73, Issue 3, pp 228-235

Tamburini M, Gangeri L, Brunelli C, Beltrami E, Boeri P, Borreani C et al (2000) Assessment of hospitalised cancer patients' needs by the Needs Evaluation Questionnaire. Ann Oncol 11:31–37

Tamburini M, Gangeri L, Brunelli C, Boeri P, Borreani C, Bosisio M, et al. Cancer patients' needs during hospitalisation: A quantitative and qualitative study. BMC Cancer 2003;3(12).

Soothill, K., Morris, S., Thomas, C., Harman, C. Francis, B and McIllmurray D,M, The universal, situational, and personal needs of cancer patients and their main carers, European Journal of Oncology Nursing, Volume 7, Issue 1, March 2003, Pages 5-13

Wen, K Y and Gustafson, D G, Needs Assessment for cancer patients and their families Health and Quality of Life Outcomes 2004, 2

The ChronoMedIt Temporal Medical Audit Framework: Progress and Agenda

Jim Warren, Thusitha Mabotuwana, Hong Yul Yang

Department of Computer Science / National Institute for Health Innovation
The University of Auckland
Tamaki Campus, Private Bag 92019, Auckland 1142 New Zealand

`{jim,thusitha,hongyul}@cs.auckland.ac.nz`

Abstract

Chronic condition management presents distinctive challenges for health information systems. ChronoMedIt is a computational framework for the analysis of electronic medical records (EMRs) to support audit of chronic conditional management. This paper describes progress to date, recent developments and research directions related to the extension and application of ChronoMedIt. We address application of the framework to epidemiology, to pay-for-performance schemes, and to support clinical intervention. We also describe enhancement to the ChronoMedIt architecture to make it more flexible.

Keywords: chronic disease management, clinical information systems, decision support, quality improvement, temporal analysis

1 Introduction

According to the World Health Organization, chronic disease was the leading cause of death worldwide in 2005 accounting for around 35 million (or 60%) of the 58 million deaths (WHO 2005). It has been projected that in 2015, 41 million (or 64%) out of all 64 million deaths will be due to chronic disease (Strong et al. 2005). In New Zealand (NZ), chronic disease is the leading cause for hospitalisation and accounts for 70% of health expenditure and 80% of all deaths (National Health Committee 2005). Thus, the burden of chronic disease is a growing concern worldwide and any incremental improvement in the management of chronic disease will have immense population health benefits, as well as cost savings and relief on the demands straining our health workforce.

To focus efforts at improving a process, it is vital to know how one is doing in the first place. A key concept here is that of *quality audit indicators* which are specific measures used to give an indication of the quality of the patient care delivered (Rosenbrand, Van Croonenborg and Wittenberg 2008). Quality audit indicators may serve to compare the relative merit of practice at one time period to another time period for a given healthcare delivery organisation – i.e., to track the trend in quality over time. For instance, we may hypothesize an upward trend in response to some newly introduced quality improvement effort, such as installation of electronic decision support (EDS) tools. We may also use quality audit indicators to compare between healthcare delivery organisations – e.g., to assess the relative performance of organisations using a particular EDS tool as compared to those not using it.

Chronic disease management, by its nature, introduces technical challenges associated with temporal reasoning, assessing time intervals and their relationships. Moreover, the input data from healthcare information systems is typically in terms of date-stamped events that only imply relevant intervals. For instance, a chronic condition diagnosis – say a diagnosis of diabetes mellitus type 2, or of essential hypertension – is made on a given date, but the implication is that the condition has existed from some (unknown) time earlier and that it will persist more-of-less for the life of the patient. Similarly, the most commonly recorded information about a medication is a prescription (and/or a dispensing of medication to fill a prescription). The prescription implies that the prescriber feels the medication is indicated for the patient. Its dosage instruction indicate how often it will be taken and how much at each administration (the *signatura* or 'sigs', which often include Latin frequency indications such as *bid* for twice daily). How long the patient is expected to take the medication is often only implied. This might be indicated at one level by the *coverage* of the prescription in terms of how much medication it can supply. A prescription for a bottle of 60 pills, with a one tablet *bid* sig and allowing two refills gives 90 days supply. Somewhat less temporal ambiguity is associated with observations such as a blood pressure, which represents a measurement on a patient at a given point in time. Although even here there is the assumption that the reading is representative of the temporal neighbourhood, particularly of all time until a newer reading is made.

In this paper we describe progress to date and ongoing efforts with respect to a computational framework that we have developed to compute quality audit indicators automatically from the data in health information systems (i.e., from electronic medical records, EMRs). The framework is called ChronoMedIt for Chronological (i.e., temporal) Medical Audit. ChronoMedIt is designed to leverage EMRs to support improvement in chronic condition management. The next section introduces the tools itself. The subsequent three sections describe its role in various areas of application. We then describe our efforts to improve the flexibility of the architecture and give a concluding summary.

2 The ChronoMedIt Framework

The ChronoMedIt software was developed as the PhD thesis of the second author under the principal supervision of the first author. It extends earlier work by the first author on the use of EMRs in general practice medicine to provide critique of management of hypertension in terms of a *temporal state-transition* model. The temporal state is the value at a given time of a set of monitored state variables (basically the set of medications prescribed, but possibly also relevant complicating diagnoses – such as diabetes, which complicates the treatment of hypertension). A temporal state-transition occurs when the state changes – e.g., a new class of medication is prescribed – with the theory that the moment the temporal state changes (transition) is a good moment to critique whether that change fits with clinical practice guidelines. An EDS structure is then formulated as a set of specific clinical alerts that are each attached to particular state transition. It was found that such a framework could give high-specificity (low 'false alarm' rate) alerts for antihypertensive prescribing (Gadzhanova et al. 2007).

The temporal state transition model was conceived as a way to provide alerts, i.e. immediate interaction feedback to cause a physician to reflect on the management of the patient in front of them. We reformulated the use of the model to quality audit more generally as a temporal audit reporting (TAR) model (Gaikwad, Warren and Kenealy 2007) that would indicate which, and how many, patients would receive a particular alert (i.e., failed a particular quality criterion) from among a given patient cohort on a time period of interest (hereafter, the *evaluation period*). Thus, one quality audit item in a TAR might indicate how many patients in a particular general medical practice were diagnosed with both hypertension and diabetes, and thus indicated for a particular class of antihypertensive called Angiotensin Converting Enzyme inhibitors (ACEi, see relevant clinical practice guidelines; Chobanian et al. 2003) but ran out of supply of it. In terms of the model, these are patients that made a temporal state transition from a having-ACEi state to not-having-ACEi state during the evaluation period.

The genesis of ChronoMedIt was a study wherein we worked closely with a New Zealand general medical practice that served a largely Pacific Islander case load. We developed a set of eight quality audit criteria that suited their needs and demonstrated that these could be derived reasonably accurately (in fact we observed 70% sensitivity and 70% specificity) from their EMR data (Warren et al. 2008). This quality audit report was based on a combination of Structure Query Language (SQL)
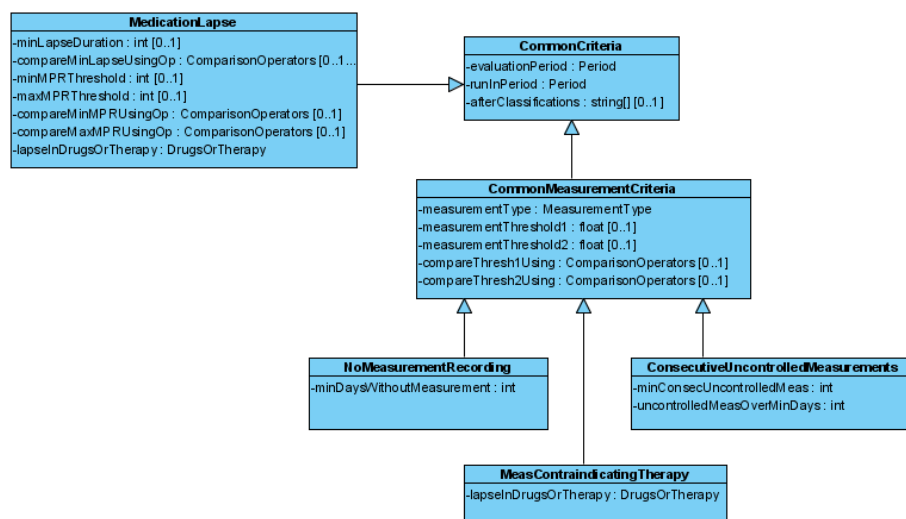

Figure 1: Criteria model for chronic condition management quality audit.

and procedural code custom to their eight criteria. ChronoMedIt was then formulated to provide a framework for answering any such queries that might arise based on abstraction of the properties of those eight.

The resultant criteria model is shown in figure 1. Any given quality audit criterion is represented as an instance of the criteria model with particular attribute values. The top level of the model is the properties that define the evaluation period and the population of interest (e.g., what 'classifications' – basically clinical diagnoses – are characteristic of the group of interest). The model then branches into two major subtypes: 1. lapses in supply of medication indicated for the population of interest; and 2. a set of subtypes related to measurements on the patient. This latter type breaks down into: (a) failure to record the outcome measure of a management activity; (b) presence of a measured value that contraindicates a therapy; and (c) sustained instances of the outcome measure being outside of acceptable bounds. We do not maintain that these are all possible quality audit criteria of interest (see section 6 for our thoughts on additional ones), merely that these are a wide family of criteria that are of interest and that circumscribe those eight that were of interest to the practice with which we were collaborating.

To make use of the criteria model required several system components:

1. A data uptake component to import data from a commercial EMR to the particular concepts of interest to our framework, including diagnostic classifications, time periods covered by medication supply (as indicated by prescriptions, or alteratively by dispensing data), observations (notably blood pressure readings) and laboratory test results. This requires interpreting medication sigs and also 'scraping' observations, such as blood pressure, out of free-text clinical notes.

2. A domain ontology to represent the values of the above concepts. This is largely a hierarchy with leaves that are the terms used in the EMR for particular diagnoses and medications organised into meaningful subtypes. For instance, sometimes we may be interested in Antihypertensive (AHT) medication generally, or

we may be interested in a sub-types (such as ACEi). As it turns out, ACEi often has the side-effect of a persistent cough, in which case another AHT class called Angiotensin Receptor Blockers (ARBs) can be substituted. Thus, we find it useful to have a mid-level concept, ACEi/ARB, to use in quality audit. Because some drug products are a combination of agents – e.g., a mixed thiazide diuretic (another AHT class) with ACEi – the ontology is not a strict hierarchy. We represent this ontology as a Web Ontology Language (OWL, McGuinness and Harmelen 2004) file and manage it with the Protégé tool (see http://protegewiki.stanford.edu/wiki/Main_Page).

3. A criteria authoring interface. We have an XML specification for criteria and a graphical user interface that interacts with the ontology to allow a user to select concepts and data value boundaries that define their quality audit requirements.

4. A query processor. The methods to assess which patients in a given EMR extract satisfy the criteria. This uses SQL with supporting stored procedures and has been validated both in terms of real-world face validity and extensive simulation testing based on comparison to results from non-SQL algorithms.

5. Data viewers. This allows the user to see the summary quality levels attained, or to see specific cases for direct follow-up. Details of the cases can be viewed as textual descriptions or as visual timeline graphs (see figure 2).

The ChronoMedIt framework has been described in detail (Mabotuwana and Warren 2010).

3 ChronoMedIt for Epidemiology

One use of ChronoMedIt is as a research tool to understand the patterns of chronic illness and how quality audit indicators align to health outcomes.

We had found there was a great deal of clinical interest in the problem of lapses in medication supply. This led us to take an interest in Medication Possession Ratio (MPR) as a measure of medication adherence (how well patients do what clinical people recommend). MPR is a 'percent of days covered' model that assesses how many days in a time period a patient had supply of their medication (Andrade et al. 2006). An MPR<80% is frequently considered as indicating a clinically significant adherence problem.

Because of the configuration of the New Zealand and Australian healthcare systems, general practitioners (GPs) have easy access to their own EMR data. This data indicates prescribing, but does not indicate dispensing. Dispensing on the other hand is best accessed in these healthcare systems via national claims reimbursement databases. These are fairly comprehensive since many medications such as AHT medications are subsidized and thus thoroughly represented in the national data. However, such collections are not designed to be timely or convenient for a local quality improvement effort (as might be initiated by a GP). This led us to investigate how well an MPR from prescribing aligned to one from national claims dispensing.

The answer turns out to be that they align reasonably well, at least for six major long-term medications that we investigated (Mabotuwana, Warren, Harrison and Kenealy 2009). For purposes of this paper, the particular answer is not so much to the point. The point is that the framework can be used to investigate questions about the
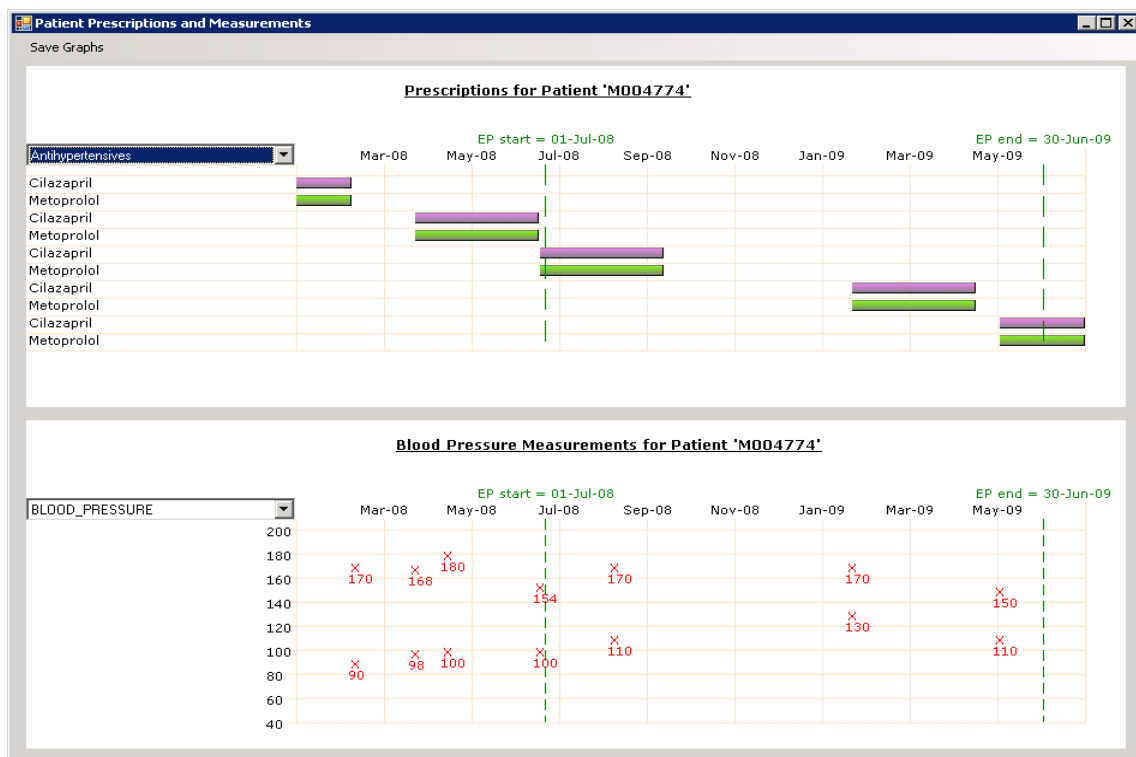
Figure 2: Timeline graph for a patient case showing a large gap in 2-agent antihypertensive therapy, an associated gap in blood pressure measurement and persistently uncontrolled blood pressure levels.

alignment of data sources. So if the GP asks, "Do I have to wait for a national e-pharmacy network, or make a request to query the national claims database in order to find patients with low MPRs?" we are able to answer that a reasonably reliable surrogate is available by looking at the EMRs from his (or her) own practice management system (PMS).

A further finding from this same study was about the alarmingly high rates of low MPR, with only half of patients achieving good adherence (MPR>=80%) on all of the six sentinel long-term medications that they were prescribed for a 15-month evaluation period and rates of good adherence to individual medications ranging from 55% to 68% (some medications appear to be more amenable to adherence than others). This class of finding is useful for characterising the prevalence of a quality issue.

A further kind of investigation is where we take the cases found through quality audit criteria and compare outcome data from those that are assessed as in violation as compared to those that are compliant. One study we did in this vein combined two of our medication supply criteria – 30 day lapse in supply, and MPR<80% – and took good adherence as having neither of these deficits present during the evaluation period. For a practice population cohort with diabetes and hypertension, good AHT adherence by this measure was associated with triple the odds of successful blood pressure control to recommended levels (Mabotuwana, Warren and Kennelly 2009). With the dangers of poor blood pressure control already well-established, this formulated a link from MPR (and, in fact, MPR based on prescribing, not dispensing) to health outcome.

Studies such as these can help to rationalise the prioritisation of health improvement efforts. Not only are health dollars limited, but the capacity of the health workforce to absorb procedural change and take on additional tasks is limited. Thus we want to establish that a problem is prevalent and important before trying to formulate and implement a relevant quality improvement strategy. In the case of medication adherence, we believe the problem to be under-prioritised and aim to continue to create more results indicating massive medication undersupply and lapse issues and their correlation to poor health outcome.

One area that is promising for further investigation in this regard is statin use – these are medication to control blood cholesterol levels. They serve both to prevent an initial cardiac disease event as well as being indicated for use by everyone who has had a previous event (e.g., a 'heart attack') to reduce the odds of another. These two cases divide into 'primary' and 'secondary' prevention. We plan to use national reimbursement data to characterise the differences in lapse in these two cases. The clinical intuition is that 6 months to a year is a danger time for adherence to statins in secondary use (i.e., it is easy to get people to adherence just after an event, but then they 'fall off the wagon'). We will use ChronoMedIt to assess the distribution of time until first 30-day-lapse and provide a quantitative basis for where to most urgently target statin adherence promotion interventions.

4 ChronoMedIt and Pay-for-Performance

A further domain of application for ChronoMedIt is potentially in performance assessment. One particular management strategy is the notion of 'pay-for-performance' – i.e., to provide financial reward for meeting quality audit targets (or, to take a 'glass half empty' view, this may be seen as financial penalties for failure to meet targets). This has been done on a massive scale in the UK through the English National Health Service's Quality and Outcomes Framework (QOF, BMA and NHS Employers 2009).

The QOF was first introduced in April 2004 as part of the General Medical Services contract. It provides a set of clinical indicators across four domains (clinical, organisational, additional services and patient experience) designed around best practice to improve the quality of service provided to patients. Each indicator is allocated a number of points and GP practices are awarded points according to how well they have performed, with associated monetary compensation. The indicators are updated annually.

Most of the widely used quality indicators use the presence of a single point-in-time measurement to determine whether a given indicator is satisfied. For example, an important QOF indicator related to the ongoing management of patients with hypertension, the one with the highest point allocation (57 QOF points) in the 'clinical' domain, is 'BP5' – "The percentage of patients with hypertension in whom the last blood pressure (measured in the previous nine months) is 150/90 or less."

It should be unsurprising to the reader to hear that the QOF is controversial. Our first attempts to participate in the debate have been to examine point-in-time QOF measures such as BP5 and look at whether patients satisfying BP5 also satisfy quality measures that assess whole time intervals, such as those based on MPR, and are free from measurement violations, such as repeated persistently high blood pressure over a time interval or long periods of missing blood pressure measurement. Using New Zealand data (which – and this could be seen as an advantage or a disadvantage – is not subject to QOF incentives) we find that there are many cases where the last blood pressure satisfied BP5, but the 'journey' was suboptimal. We find long periods without measurement to be more the problem than persistently high measures (perhaps if the patient were measured they would have measured high) (Mabotuwana et al. 2010). At any rate, it supports an argument that a system incentivizing just the last measure, say annually, coming good leaves ample room for suboptimal management when defined in more depth over the entire time period.

More broadly speaking, this illustrates use of ChronoMedIt to inform the debate over what performance indicators to incentivize. Further work in this direction could either be look at how ChronoMedIt could be used by practices to identify care process deficits that anticipate failures to meet pay-for-performance targets, or to fuel debate that the incentive systems themselves should be reformed.

5   ChronoMedIt for Clinical Intervention

Up to this point we have mainly spoke of ChronoMedIt as a tool to measure a whole cohort of patients and hence provide an assessment of a health provider setting (whether it is a single general practice or a whole country). The alternative use of the same underlying query results, however, is for 'case finding.' That is, rather than being interested in the percentage of cases in violation of a criterion, to be interested in just who those individuals are so as to follow up and improve the situation. The interactive EDS situation is like this for a cohort of size one – if there is a violation, then the user gets an alert. Our technology could be integrated with health information systems for this use, however, we are focusing instead on a process that might be done, say, quarterly where a list of cases is established and then systematically pursued for improvement. It is worth saying that we believe both cases are valid and useful (and we return to this point at the end of this section).

We have investigated the direct clinical use of ChronoMedIt reporting in New Zealand Health Research Council (HRC) Feasibility Study, called Adherence Innovation in Medication use for Health Improvement (AIM-HI). The AIM-HI intervention has involved the identification of a cohort of patients (approximately 200) who have an overall AHT MPR<80% for a 12-month period. These patient are then being actively followed up by practice nurses for a subsequent 12 months with the aim of improving their AHT adherence and their blood pressure control. As a Feasibility Study the size of the experiment is not designed to establish a statistically significant outcome, but rather to determine that the intervention is practical.

The AIM-HI assessment has included patient and health provider focus groups. Some of the qualitative findings include that patients are mostly happy with a telephone call to remind them when it is time to return to the practice. In our group at least, the option of a cell phone reminder (possibly an automated SMS message) was questionable. One should not be lulled by statistics indicating high levels of cell phone ownership – among our participants, among those around age 60, it was raised that they do own a cell phone, but they don't necessarily keep it charged. Reactions to home visits or postal reminders were variable. A difficult issue was the timing of follow-up, however, as individuals differed in how much warning they wanted. Calling and expecting that they could come in one just a couple days notice was considered rude and disrespectful of how busy their lives were otherwise; but calling a week or more in advance risked heading off the opportunity for them to remember to make the appointment for themselves.

Staff issues included dissatisfaction with identifying low-MPR patients for special follow-up when some other patients may be sicker. Also, there is the general issue of a hard-pressed health workforce. There is a belief that human follow-up is the best way to get to know individual adherence barriers, negotiate around the barriers and influence the patients to do better. But, of course, this requires time that could be used in other activities.

We are pursuing two distinct agenda to follow on from the AIM-HI Feasibility Study. One is around automated technologies that largely avoid the health workforce issue. A particularly relevant automated intervention has been STOMP (Stop Smoking Over Mobile Phone) which demonstrated significant success in smoking quite rates for Maori and non-Maori participants. This inspires the notion that we may be able to engage patients through a programme of SMS messaging to improve their medication adherence behaviour. Obviously this is limited to the cell phone using segment of the cohort. For others it may be that the cornerstone should be something home based that monitors the actual medication administration events. There is also the possibility of using Web-based education, but with a focus on younger family members to influence the patient, rather than aimed directly (or at least exclusively) at the patient themselves.

We are also looking at further development of health workforce based interventions. We are proposing a trial where medical students doing training placements in general practice help patients identified via ChronoMedIt to work through adherence issues. The immediate outcome (even before MPR or blood pressure improvement) would be to see a greater level of Shared Decision Making (Dy 2007) in doctor-patient consults, indicating that the patients were taking the decision be on medication seriously.

These cases have illustrated use of ChronoMedIt reporting directly in clinical care. Health Informatics has focused on great deal on EDS technology. For instance, the ATHENA system is an excellent example of a comprehensive decision support system for hypertension management (Goldstein et al. 2000). Also, in New Zealand, there is substantial uptake of the PREDICT system which assess risk of cardiovascular disease events and then provides tailored recommendations of how to manage down that risk (Kerr et al. 2008). PREDICT has now had over 100,000 sessions of use with patients in New Zealand. We believe that the use of a case finding approach is complementary to these EDS efforts. With a list of cases, all with a common problem, there is the potential to assign a worker to that particular task. That worker will very often not need to be a physician, and may not even need a great deal of education. Furthermore, the two approaches can link in that one action from systematic follow-up of a case by a non-physician could be to tag that case in the practice management system so that the GP is alerted to the issue next time they see the patient.

6   Enhancing ChronoMedIt

We have been working to enhance the ChronoMedIt software for greater flexibility. Key elements of these efforts involve: the data uptake mechanisms; the internal representation of cases; and the family of audit criteria supported.

With respect to the data uptake mechanism, we had initially focused only on the market-leading New Zealand General Practice management system, Medtech32 by Medtech Global, and also national dispensing data as an auxiliary source. Obviously this is limiting, even within New Zealand, but also of course internationally.

To generalise the data uptake mechanism we have made the ChronoMedIt architecture more modular (see figure 3). There will always be some elements that will need to be specific to a given source EMR, but there are a couple components we believe will be quite general. One is notes 'scraping' for elements such as blood pressure (which, although they can be represented formally as observation data in Medtech32, are often just typed in as text in-line with other notes). The ways that blood pressures and other observations are typed into free-text notes are likely to be similar irrespective of the software used. Thus it is worth abstracting this module from the system-specific data uptake. We have done this through an XML based description of extraction pattern matching rules and an associated component that applies the rules to the appropriate files as part of the PMS-specific data uptake adapter. A similar generality holds for medication sigs – these are essentially universal, not system-specific, and hence the interpreter for them forms a valuable resource that can be reused as we tailor data uptake adapters to particular PMS/EMR solutions.

Different systems (and different jurisdictions) will use different terminology sets for diagnoses and, to a lesser extent, medications. These variations are readily managed through our ontology, which allows multiple sets of terminology to be managed under higher-level relatively universal concepts. For example, New Zealand uses Read Clinical Codes for diagnosis coding in general practices, whereas International Classification of Primary Care (ICPC) codes are more common in Australia; while the details will differ, both systems will have term-code pairs that cluster under a high-level concept like Hypertension.

Ideally, we would like a standards based solution such that EMR systems could produce data ready for use by ChronoMedIt without system-specific adaptation. This level of interoperability is probably infeasible in the near term. We see in the Health Level 7 Clinical Document Architecture (CDA) (Dolin et al. 2006), however, the promise least for a framework to allow such interoperability. We aim to define a CDA input format that provides the data required for ChronoMedIt analysis. With such a format specification, in theory an EMR user with a CDA-compliant system could define a reporting format that met our input requirements directly.

With respect to internal representation, we have unified the concepts of observations and laboratory test results in the internal representation. These items come in very differently in many source EMRs. In Medtech32 the blood pressures are largely recorded in free-text notes. Laboratory test results come in over a third-party network as Health Level 7 messages and are stored in an inbox for the physician who ordered them. Despite the initial differences in source, we have come to realise that the role of each is equivalent in our query criteria – thus, a cholesterol level measured in an external laboratory or a blood pressure measured at the doctor's office might equally be a measure that is, for instance, not taken as often as appropriate, or is hovering repeatedly at an unacceptable level.

Further complexity is introduced by derivative and composite measures. Blood pressure itself of course is two numbers, the systolic and diastolic' cholesterol results are typically presented as four numbers. A quality criterion may refer to ranges on multiple numbers (e.g., a blood pressure under 140/90, requiring both systolic and diastolic to be under their respective targets) or as a function of numbers (as in an HDL:LDL, "good cholesterol" to "bad cholesterol" ratio). We have not reached a comprehensive solution for these cases.

An open-ended challenge is with respect to extending the types of queries managed by the system. MPR was an early extension of ChronoMedIt – it was not directly indicated by the initial eight criteria, but rather emerged from the pharmacoepidemiology literature as we researched the medication adherence problems that were so prevalent as initially indicated by simple lapses in medication supply. We have recently identified another query area that may attract clinical interest with respect to therapy that remains static despite failure to meet target. This still fits the larger criterion object model, but will require an extension to the internal data model because 'static' therapy implies a lack of dose intensification. So far we have not been concerned with dose levels beyond their role in determining the duration of medication supply.

7 Conclusion

We have developed a framework for analysis of EMRs to provide quality audit reporting in chronic condition management. Our ChronoMedIt framework provides software for data uptake from EMRs, management of relevant concept ontologies, user specification of quality audit criteria details, query processing per se, and reporting, including case timeline visualisation.

In terms of related work by others, our work most closely resembles the IDAN/KNAVE II framework (Boaz and Shahar 2005; Shahar et al. 2006) for temporal abstraction on clinical data. Our main point of distinction is actually with respect to our *lack* of generality; that is, ChronoMedIt is not designed to allow users to explore arbitrary temporal relationships in clinical data. Rather, it is focused on management of major classes of temporal query that we have found to be relevant for quality audit of chronic condition management.
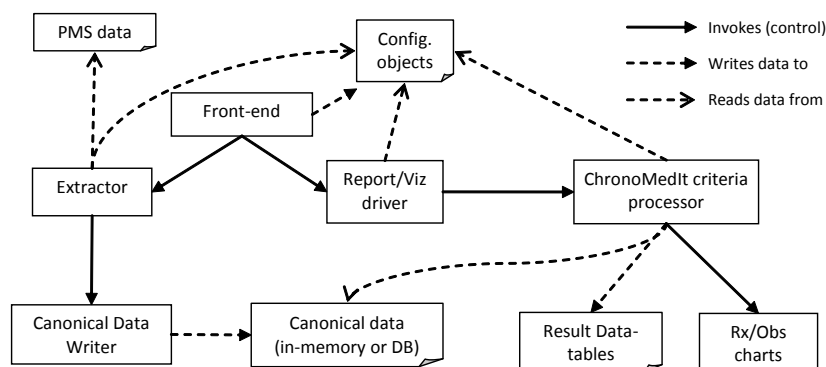

Figure 3: ChronoMedIt architecture.

We have identified applications of ChronoMedIt in epidemiological research, performance management and clinical intervention. We are working to make the framework more flexible for wider use. We believe we have demonstrated the technical ability to make meaningful measurements of the quality of healthcare processes, and to identify patient cases with care that is suboptimal in well-defined ways. A larger question, however, lies in how best to utilise such capability to enhance healthcare delivery. We are interested in collaboration to apply the framework more broadly, especially where it may link to sustainable health improvement.

8   Acknowledgments

9   References

Andrade, S.E., Kahler, K.H., Frech, F., and Chan, K.A. (2006): Methods for evaluation of medication adherence and persistence using automated databases. *Pharmacoepidemiology and Drug Safety*, 15(8): pp. 565-74; discussion 575-7.

Boaz, D. and Shahar, Y. (2005): A framework for distributed mediation of temporal-abstraction queries to clinical databases. *Artif Intell Med*, 34(1): pp. 3-24.

BMA and NHS Employers (2009): Quality and Outcomes Framework guidance for GMS contract 2009/10: Delivering investment in general practice. http://www.bma.org.uk/images/qof0309_tcm41-184025.pdf Accessed 29 September, 2010.

Chobanian, A.V., Bakris, G.L., Black, H.R., Cushman, W.C., Green, L.A., Izzo, J.L., Jr., Jones, D.W., Materson, B.J., Oparil, S., Wright, J.T., Jr., and Roccella, E.J. (2003): Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension*, 42(6): pp. 1206-52.

Dolin, R.H., Alschuler, L., Boyer, S., Beebe, C., Behlen, F.M., Biron, P.V., and Shabo Shvo, A. (2006): HL7 Clinical Document Architecture, Release 2. *Journal of the American Medical Informatics Association*, 13(1): pp. 30-9.

Dy, S.M. (2007): Instruments for evaluating shared medical decision making: a structured literature review. *Med Care Res Rev*, 64(6): 623-49.

Gadzhanova, S., Iankov, II, Warren, J.R., Stanek, J., Misan, G.M., Baig, Z., and Ponte, L. (2007): Developing high-specificity anti-hypertensive alerts by therapeutic state analysis of electronic prescribing records. *Journal of the American Medical Informatics Association*, 14(1): pp. 100-9.

Gaikwad, R., Warren, J., and Kenealy, T. (2007): The TAR model: use of therapeutic state transitions for quality assurance reporting in chronic disease management. *Studies in Health Technology and Informatics*, 129(Pt 2): pp. 866-70.

Goldstein, M.K., Hoffman, B.B., Coleman, R.W., Musen, M.A., Tu, S.W., Advani, A., Shankar, R., and O'Connor, M. (2000): Implementing clinical practice guidelines while taking account of changing evidence: ATHENA DSS, an easily modifiable decision-support system for managing hypertension in primary care. *AMIA Annual Symposium Proceedings*, pp. 300-4.

Kerr, A.J., McLachlan, A., Furness, S., Broad, J., Riddell, T., Jackson, R., and Wells, S. (2008): The burden of modifiable cardiovascular risk factors in the coronary care unit by age, ethnicity, and socioeconomic status - PREDICT CVD-9. *New Zealand Medical Journal*, 121(1285): pp. 20-33.

Mabotuwana, T. and Warren, J. (2010): ChronoMedIt - A computational quality audit framework for better management of patients with chronic conditions. *Journal of Biomedical Informatics*, 43(1): pp. 144-58.

Mabotuwana, T., Warren, J., Elley, C.R., Kennelly, J., Paton, C., Warren, D., Chang Wai, L., and Wells, S. (2010): Use of interval based quality indicators in blood pressure management to enhance quality of pay-for-performance incentives – comparison to two indicators from the Quality and Outcomes Framework. *Quality in Primary Care*, 18(2): pp. 93-101.

Mabotuwana, T., Warren, J., Harrison, J., and Kenealy, T. (2009): What can primary care prescribing data tell us about individual adherence to long-term medication?-comparison to pharmacy dispensing data. *Pharmacoepidemiology and Drug Safety*, 18(10): pp. 956-64.

Mabotuwana, T., Warren, J., and Kennelly, J. (2009): A computational framework to identify patients with poor adherence to blood pressure lowering medication. *International Journal of Medical Informatics*, 78(11): pp. 745-56.

McGuinness, D.L. and Harmelen, F.v. (2004): OWL Web Ontology Language Overview. http://www.w3.org/TR/owl-features/ Accessed 29 September, 2010.

National Health Committee (2005): *People with chronic health conditions: a discussion paper.* Wellington: National Health Committee.

Rosenbrand, K., Van Croonenborg, J., and Wittenberg, J. (2008): Guideline development. In *Computer-based Medical Guideline and Protocols: A Primer and Current Trends*, A.T. Teije, S. Miksch, and P. Lucas (eds). IOS Press: Amsterdam. pp. 3-21.

Shahar, Y., Goren-Bar, D., Boaz, D., and Tahan, G. (2006): Distributed, intelligent, interactive visualization and exploration of time-oriented clinical data and their abstractions. *Artif Intell Med*, 38(2): pp. 115-35.

Strong, K., Mathers, C., Leeder, S., and Beaglehole, R. (2005): Preventing chronic diseases: how many lives can we save? *Lancet*, 366(9496): pp. 1578-82.

Warren, J., Gaikwad, R., Mabotuwana, T., Kennelly, J., and Kenealy, T. (2008): Utilising practice management system data for quality improvement in use of blood pressure lowering medications in general practice. *N Z Med J*, 121(1285): pp. 53-62.

WHO (2005): *Preventing chronic diseases: a vital investment*. Geneva: World Health Organization.

Leveraging Web 2.0 and Consumer Devices for Improving Elderlies' Health

Jaspaljeet Singh Dhillion   Christof Lutteroth   Burkhard C. Wünsche

Department of Computer Science
University of Auckland
Private Bag 92019, Auckland, New Zealand

jran055@aucklanduni.ac.nz; lutteroth@cs.auckland.ac.nz; burkhard@cs.auckland.ac.nz

Abstract

With a growing elderly population in many developed countries, technologies for supporting elderly healthcare are becoming more and more important. As technologies such as the Web and consumer level devices evolve, this opens up new opportunities for telehealthcare. In this paper, we discuss how the paradigm known as the Web 2.0 can be leveraged to develop solutions that are more patient-centric and empower the patients, especially the elderly, to manage their own health from home. We critically analyse popular Web 2.0 health applications and propose a framework for overcoming their shortcomings. Furthermore, we describe how consumer level devices such as off-the-shelf computers and motion sensing input devices can be used to make telehealthcare more accessible and affordable. Devices such as iPhones and the Wiimote controllers can enable patients to perform rehabilitation and prevention activities such as exercises at home. We argue that by combining the potentials of Web 2.0 and new consumer devices, a more holistic approach to telehealthcare can be achieved. ·

Keywords: patient networks, motion input, telehealth

1   Introduction

The web is becoming the ultimate resource for health information. Evidence from Elkin (2008) suggests that health consumers are increasingly depending on the Web to look for health information. The survey also indicates that most people (75% of the respondents) look for information about a specific condition, disease or symptom. Likewise, Fox (2006) found that 80% of adult Internet users in America (about 113 million people) have searched for health information in 2006. Besides this, people are increasingly moving beyond passive searches to interactive forums, blogs, and other social media. This trend is often referred to as Web 2.0.

Web 2.0 is commonly known as "the web as a platform" (O'Reilly 2005, Anderson 2007) and is associated with web applications and services that facilitate interactive information sharing, rich user experience, dynamic content and user-centred design.

Increasingly, more websites are built upon these ideas and are transforming the Web from a simple place to store information to a dynamic place where people regularly gather and interact. Examples of Web 2.0 include social networking sites, blogs, wikis, multimedia sharing sites, hosted services, web applications, mashups and folksonomies. The use of Web 2.0 in healthcare is evolving as more applications and services targeting health professionals and patients are being developed. With that trend, the term Health 2.0 (also referred as Medicine 2.0) is becoming popular.

One of the most significant and current discussions in the healthcare sector is about empowering patients to manage, i.e. maintain or improve, their own health. There has been significant interest of researchers and system designers to develop patient-focussed health systems oriented towards this goal. Some examples of Web 2.0 applications include PatientsLikeMe, CureTogether, DailyStrength, Disaboom, SugarStas, and DailyBurn. These applications provide direct patient support, promote disease awareness, and encourage positive and proactive behaviours to stay healthy. According to Sarasohn-Kahn (2008), online peer support is also attractive to older adults and those living with chronic conditions, particularly if it is delivered on technology that is familiar and convenient.

Likewise, telehealth and telecare solutions are rapidly gaining in popularity because of their promise to use existing health care resources more effectively. They aim to lower costs and enable health consumers to transmit health data to clinicians from home. Onor et al. (2008) indicates that in one of the primary target groups, the elderly, the satisfaction with telecare is significantly higher compared to daycare centers or nursing homes. However, as Dishman (2004) points out, there are fundamental concerns that telehealth systems are mostly focussed on treating disease, rather than preventing it. Many existing telehealth systems, as analysed by Singh et al. (2010a), suffer from high initial costs, cannot be extended by third parties, and require extra costs to add new functionalities. They also do not address the social and psychological needs of the patient.

Recently, Elder Brief (2010) reported on an easy-to-use emotional networking platform developed by Emota, which uses touch screen Internet tablets and picture frames in a cartoon-like format to help seniors maintain closeness and mutual awareness with their loved ones. This product is aimed to tackle isolation, which is a huge issue for elderly people living alone. However,

products like this one still exhibit many of the disadvantages pointed out above.

Web 2.0 health applications and services are rapidly gaining attention from patients and professionals, as they have the potential to extend traditional healthcare delivery models, empower patient self-care and provide social support. The increasing numbers of registered members of healthcare websites indicate that patients are starting to manage their health online: independently, with peers, with online affinity communities and with medical professionals. According to Swan (2009), most patient-focused social health networks offer the basic level of service, emotional support and information sharing, for a variety of medical conditions. Most Web 2.0 health applications and services have an emphasis on communication, information sharing and community. The Web has evolved from supporting health information seeking to support health-motivated social communication, and it is time to offer also health applications that support proactive healthcare, i.e. that actively engage users in healthcare activities.

Instead of expecting health consumers to purchase expensive devices to engage them in their own healthcare, we claim that it is possible to provide affordable healthcare to consumers by leveraging mainstream consumer devices. Such devices include personal computers, webcams, iPhones and Wii controllers. Consumers are more likely to use devices they already have with them or in their home to take care of their health, rather than buying special hardware. Furthermore, we claim that the integration of Web 2.0 healthcare approaches with device-supported approaches for proactive healthcare is a crucial step to make telehealth more accessible and more effective for elderly.

To support these statements, we evaluate existing popular Web 2.0 health applications from a patient perspective and discuss their strengths and weaknesses. Based on the results we propose a novel framework for a ubiquitous patient-centric, web-based telehealth system, which is designed to be unobtrusive and enable patients to improve their own health. We believe that such a system will be an important contribution to patient networks, especially for the elderly, to improve health and to address social isolation. Furthermore, we discuss how consumer devices can be leveraged to enable dynamic interaction and exercise for elderly users.

Section 2 motivates the need of Web 2.0 applications that are specifically designed for the elderly. Section 3 reviews and analyses Web 2.0 health applications, which are grouped into 3 categories: health information interchange, social networking, and health monitoring. Section 4 discusses shortcomings of these applications with regard to telehealth for the elderly. We introduce our own approach for building an elderly-freindly telehealth system in Section 5. We discuss in this section how we can leverage Web 2.0 and consumer devices for developing health applications, rather than building a system from scratch. We conclude the paper in Section 6.

2 Motivation

The demographics of frequent Internet surfers vary from region to region. Usually they are assumed to be young,

college-educated, middle class, suburban living individuals. However, this perceived demographic description fails to address a section of the population, the elderly, who have a significant impact on online media and social networking. Elderly are usually classified as the age group that generally covers individuals from 65 years upward.

The impact of elderly on the media is somewhat significant, but often does not get the amount of attention that is justly deserved. In a recent report in Generations Online 2009 by the Pew Internet & American Life Project, Jones (2009) states that the fastest growth in Internet use is being driven by the older age groups, starting at 55. Interestingly, the 70-75 year-old age group has increased its Internet use more than any other group since 2005. While just over one-fourth (26%) of 70-75 year olds were online in 2005, 45% of that age group are currently online. Increasingly, elderly users are searching for health information over the web. According to Jones (2009), researching health information is the third most popular online activity with the most senior age group, after email and online search.

However, current research shows that the elderly are not utilizing existing social networking websites such as Facebook and MySpace, as compared to the younger group of users. Findings of Jones (2009) highlight that instant messaging, social networking, and blogging have gained ground as communications tools, but email remains the most popular online activity, particularly among older internet users. Survey results of Fox (2009) show that only 7% of adults 65 and older have posted profiles on online social networks. Statistics from Inside Facebook, by Smith (2010) clearly show that the 55-65 year-old age group represents the smallest group of Facebook users, only 7% of the total user population. Facebook crosses 60 million monthly American users, but according to Smith (2009), they seem to face challenges in retaining older users as fewer users of age over 55 are returning to the site. Most of the Web 2.0 health applications are created for a general audience, meaning that they not designed with elderly in mind.

3 Health 2.0 Today

There is a variety of Web 2.0 patient-focused health applications and services available to health consumers. Some of them successfully empower, engage, and educate their users. As the large and growing number of registered members illustrates, such systems are gaining popularity. As a result, it is importat to look at them and reflect on the services and benefits they provide.

This section analyses some of the most popular web health applications, dividing them into the following groups: applications for health information interchange, which help patients to find health-related information; social networking applications that help patients to communicate with other patients suffering from similar conditions; and health monitoring applications that help users to keep track of their own health.

3.1 Health Information Interchange

One of the characteristics of Web 2.0 applications is the ability to leverage the collective knowledge of its users. Applications for health information interchange do so

with the aim of providing disease-related information, in most cases directly from the people concerned.

These applications are mainly designed to educate patients on various diseases, symptoms and treatments via other patients that have experienced the same health condition. Conditions are often presented with easy-to-understand graphs or charts, which are based on health data aggregated from many patients. Some websites also produce good reports and summaries for patients. They frequently enable patients to connect with other patients, e.g. by posting comments to other patients' health data.

3.1.1 PatientsLikeMe

According to Domingo (2010), PatientsLikeMe is a social networking website with more than 58,000 registered patients as of March 2010. Patients are organised by disease, and share information about treatments and symptoms of various life-alerting diseases, experience with drug side effects, and how to improve their lives. Patients' health records are made available to other users on the site, which raises privacy concerns.

There are free tools for patients to track their medications, symptoms, and health outcomes. The health information is presented with easy-to-read charts that enable members to search for medical profiles that best match their own. The site empowers patients to compare their health progress, treatments and dosage levels with other patients on the network. The site makes money by selling the data to drug companies.

PatientsLikeMe also allows patients to help other patients to deal with their diseases. Patients, especially those experiencing a life threatening disease, can freely and honestly share their sufferings on the website, and other patients can reply and provide emotional support. Hence the site supports lonely users, by making them feel less isolated. Users can also print a "doctor sheet" that summarizes their health progress. However, the main function of the site is that of information aggregation.

3.1.2 CureTogether

According to Domingo (2010), CureTogether is a community based health site whereby patients and researchers come together to share information and help to find cures for some of the most painful, prevalent, and chronic conditions. As of as of March 2010, it has 7,200 registered patients. CureTogether helps people to anonymously track and compare health data, to better understand their bodies, make more informed treatment decisions, and contribute data to research. It started with three conditions (Migraine, Endometriosis and Vulvodynia), and according to Carmichael (2009), it was expanded to support up to 228 diseases as people wrote in to request that their conditions should be added to this ongoing study. It enables patients to connect with other patients that most closely match their health conditions. Registered patients get access to condition reports. The health information is presented with easy-to-understand charts, which are systematically categorised as symptoms, treatments, causes and related conditions.

3.2 Social Networking

Whereas the sites described in the last section focus on health information, the ones described in this section have as a focus the social interaction between their users. Social networking applications provide users with emotional support to cope with their health condition. They allow patients, especially elderly and disabled people, to share their pain and socialise with others on the network. Users can mostly communicate via mail, posting comments and chatting. For house-bound patients, the network provides a social life they might otherwise not have.

3.2.1 DailyStrength

DailyStrength is a social network focused on helping people overcome health and life challenges through advice and friendly support. Over 500 support groups organized around specific diseases are hosted on this web site. Users can find detailed information on treatments and symptoms, in addition to professional and reliable health advice. Users can also describe their experiences and share the successes or failures they have had with different treatments. Tools to set goals and vote on the effectiveness of treatments are available. On the main page users find a list of popular support groups, the latest activity on the website, the latest topics in the "ask an expert" section, a Facebook section, and a DailyStrength store section. The website organizes its content under the main categories of support groups, expert advice, treatment, and people. Users can find links on the bottom of the webpage that target specific issues.

One of the most popular features on the site is a "virtual hug". A member's profile page will keep track of all their friends, messages from other members, their journal entries and all the hugs they receive in their "hug book". The website also contains a chat bar that allows users to speak to each other in real time and perform other tasks.

3.2.2 Disaboom

Disaboom is a social networking website dedicated to improving the way individuals with disabilities live their lives. As of March 2010, it has about 100,000 members. Disaboom gives handicapped individuals a space to make friends, and discuss day-to-day challenges. Furthermore, it helps to locate trustworthy sources for all types of life issues, ranging from rehabilitation advice and pain management to finding jobs or enriching family relationships. Members are provided with tools to blog, participate in discussion boards, and engage in real-time chat.

Domingo (2010) reports that Disaboom is vital for users who are not able to leave their homes or feel alone in their health struggles. Users, while remaining at home, are able to watch movies with one another, chatting online while watching.

3.3 Health Monitoring

Health monitoring sites help users to keep track of health-related parameters, visualize them, and share that data with others. They empower patients to track any progress towards their health and fitness goals. Most of the applications allow access to and input of health data from anywhere through mobile devices.

3.3.1 SugarStats

SugarStats is a web-based blood-sugar tracker and diabetes management system. Users are provided with a simple interface to track, monitor and access their glucose levels and diabetic statistics to spot dangerous trends and better manage their diabetic health. SugarStats enables users to input and access their data via mobile phone, email and Twitter. Users can track blood sugar glucose levels along with the elements that effect those levels such as medication, food intake and physical activity. They can then share this information with health care professionals, family and friends to get further support and advice.

3.3.2 DailyBurn

DailyBurn is an online application dedicated to helping people reach their fitness goals. According to Wauters (2010), it has about 500,000 members. DailyBurn helps users to lose weight and gain muscle by tracking their workouts and the food they consume. The site offers a database of foods, allowing users to quickly figure out how many calories they have eaten throughout the day. Users can input data about their daily exercise regime (number of miles run, bench-press weight, etc.), so they can track their progress over time using graphs. DailyBurn also offers an iPhone application that allow users to track and update their progress through their phone.

Apart from tracking fitness goals such as pounds lost and changes in resting heart rate, users can participate in DailyBurn's social network. Group-designed workouts and challenges give them fresh and interesting goals. Users can plan meals, and keep track of successes and problems in a training journal. Users can find a partner to motivate them, track their results.

4 Web 2.0 and the Elderly

The emphasis of Web 2.0 health applications and services has been on communication, information sharing and community, rather than tackling the bigger challenges such as providing medical diagnosis or treating disease over the web. The latter is much more difficult because it usually requires input from certified experts. However, early diagnosis and treatment of health conditions is a crucial element for elderly people's healthcare.

The applications discussed above are created with a single objective in mind: either to educate patients, to provide social support or to monitor health parameters. None of the popular applications provides all three types of functionalities. However, for a system to be truly effective and have a holistic impact on health, it needs to offer most of the important functionalities such as education, monitoring, diagnosis, rehabilitation and social support. For elderly people, it is often not feasible to work with many different systems, and a single integrated user interface is necessary (Singh et al, 2010a).

Furthermore, the health applications described above are created for people who are comfortable with computers, but they are not suitable for elderly who have special usability requirements. However, elderly are disproportionately more affected with chronic health problems, physical and mental challenges, and loneliness.

They could benefit tremendously from the Web 2.0, if it could be made more accessible to them.

The popular websites that encourage a proactive approach to prevent disease, maintain or improve health such as DailyBurn, are not suitable for elderly. General exercises are often too hard for elderly people, and typical health conditions of elderly people require special types of exercises. For instance, patients suffering from Parkinson disease need support for performing hand exercises to improve the mobility of their hands. Active participation such as computer-supported exercise is an important function of elderly healthcare, which is not well supported in today's information-centric Web 2.0.

5 A Telehealth Framework for the Elderly

Based on the shortcomings indentified in the previous section, we propose a framework for a novel web-based telehealth system, *Healthcare4Life*, which is illustrated in Figure 1.
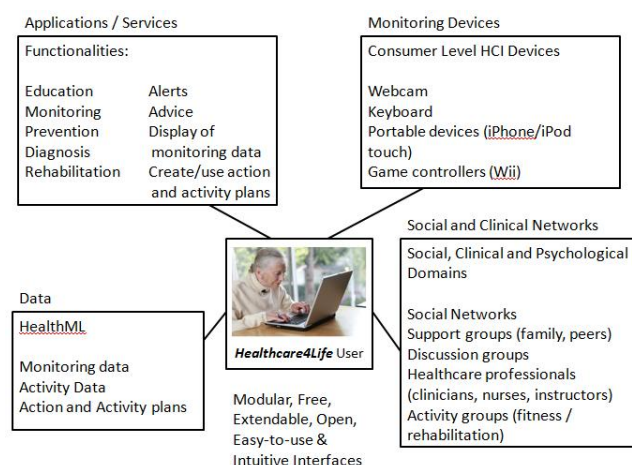
Figure 1: Framework for Healthcare4Life (Singh et al, 2010a)

Healthcare4life is a web-based framework that combines the power of social networking with telehealth systems in empowering patients, especially the elderly, to manage their health independently from home. It addresses the restrictions of traditional telehealth systems, and aims at making telehealth for the elderly more widely available, affordable and extendable. Similar to existing health 2.0 systems, it tries to encourage positive lifestyle change by letting elderlies manage their own healthcare goals.

Healthcare4Life promotes social networks over clinical networks to motivate users, especially the elderly, to take control of their health and to address social isolation. Similar to the social networking applications discussed above, it serves as a medium of interaction between patients. Patients suffering from the same disease are able to share experiences, and most importantly, do exercise together via a video link or in a virtual environment.
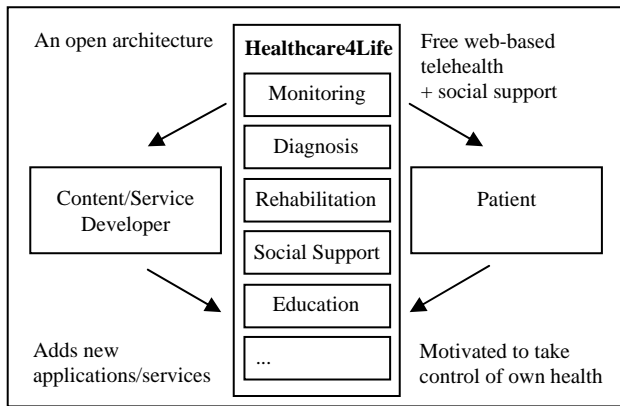
Figure 2: Overview of how Healthcare4Life functions on the surface level

Similar to Facebook, the framework has an open architecture that enables third-party providers to add new content and functionalities (see Figure 2). Applications added into Healthcare4Life fall into the following categories: monitoring, diagnosis, education, rehabilitation and social support. Developers can design applications for these categories in the form of serious games, interactive web pages and expert systems. Unlike existing social networking websites that mainly focus on 'just for fun' or advertisement-based applications, it aims at attracting serious developers to build and share health related applications with patients. Developers get useful feedback from potential users through a ranking system, which displays user satisfaction and the popularity of each service.

To keep overheads low, the framework supports the usage of common consumer level HCI devices such as webcam, keyboard, iPhone, and Wii remote controller with the system. To date, we have developed two iPhone games (discussed in section 5.3) that enable users to perform rehabilitation exercises for the elbow and shoulder joints.

The Healthcare4Life framework can be used in developing extendable ubiquitous patient-centric systems. Currently, the framework is being employed to implement a web-based telehealth system. The system is currently at a prototyping stage. A systematic assessment will be conducted to test the system at elderly community centres using interviews, system usage data, and questionnaires. A summative evaluation will be executed, i.e. evaluation will be done after the system has been developed completely.

Since our system targets the elderlies, usability testing is crucial to ensure that the system meets their expectations (Singh et al, 2010b). The evaluation will mainly focus on two aspects of user experience: performance and satisfaction. User performance is about measuring the ability of the user to successfully accomplish a set of planned tasks (Tullis & Albert, 2008). The user performance will be evaluated by measuring how hard it is for elderly to perform typical tasks in the system, e.g. how long they take to access health information. This is done through session recording, which can be integrated into a web-based system such as ours relatively easily (Weber 2008). User satisfaction refers to how the user feels about the interaction with the system (e.g. if it is easy to use, or confusing).

5.1 Leveraging the Web 2.0

Weitzel et al. (2009) have illustrated the use of Web 2.0 technology in providing extended care networks for chronic disease management and elderly care. Recently, Weitzel et al. (2010) have presented a Web 2.0 model for patient-centred health informatics applications. The suggested model uses open technologies such as OpenSocial, REST, and Open Authentication. OpenSocial is an emerging standard that defines a component model for browsers, called gadgets, as well as APIs for accessing a person's profile, activities, and relationships, such as friends. The model has good potential in creating a standards-based platform to create rich, highly interactive, community-oriented websites that can easily embed third-party applications. This model is used for Healthcare4Life, in order to create a platform that has the ability to integrate with other social networking websites (such as Orkut and MySpace) and enables developers to add content and functionalities into the system.

It is important to realize that it is extremely hard to create an entirely new social network that satisfies the needs of the elderly. However, it is possible to leverage existing networks and adapt them so that they become accessible to the elderly. Recently, Ko et al. (2010) reported about the connect services major social-networking sites have launched, e.g. Facebook platform, Google Friend Connect, and MySpaceID. They enable third-party sites to develop social applications and extend their services without having to host or build their own social network. These extensions allow third-party sites to leverage the social-networking site's features.

Jaffe (2010) describes tools such as Facebook open graph and others, called social plugins for third-party websites, which enable web developers to extend the social experience of visitors on their site. For example, third-party sites can exploit the authentication services provided by Facebook so that users need not create another username and password to access the third-party site. That is, the users can draw on their credentials from Facebook and their established profile.

Two other promising plugins include an Activity Feed, which gives users live updates of their Facebook friends' activities on that site, and a Live Stream, which acts as a Facebook "wall" embedded on the site and allows for users to comment and post discussion pieces during a live event to the website.

Open Graph, according to Axon (2010), will allow Facebook and participating websites to use social information to customize each visiting experience. For example, if the user has stated a preference on Facebook for a certain kind of music or food, another site can use that information to highlight certain content. Hence, by intergrating Facebook's social features Healthcare4Life will enable elderly to connect with their family, especially grandchildren that have Facebook accounts.

Rather than trying to recreate large amounts of existing functionality, the functions of other Web 2.0 applications can be leveraged as well. Such applications are could be useful for elderly if they were adapted to their user interface and specific content requirements. For example,

DailyBurn offers a public API, allowing third-parties to integrate with this website. Rather than tracking burnt calories, this functionality could also be used to track range-of-motion, or other parameters that are more relevant for elderlies. For real-time communication, the Skype API provides a mechanism for 3rd-party scripts and applications to control Skype UI functions, and implement additional or improved features to complement Skype. The API has two layers: The Communication Layer is a set of methods for external applications to establish a connection to a Skype client and communicate with it. The Command Protocol Layer is a text-based "language" that external applications can use to speak to a Skype client, once communication channel is established by Communication Layer. Although APIs that are not entirely web-based are harder to integrate, there are still possibilities to do so, e.g. with the help of browser plug-ins.

5.2   Leveraging Mainstream Consumer Devices

The recent proliferation of motion sensing devices such as iPhone and Wii Remote makes it possible to leverage such devices for healthcare. They can be used to allow patients to perform guided rehabilitation activities to improve their condition, as part of a telehealth system. These devices contain motion sensors such as accelerometers that can measure positions, velocity and direction vectors. This opens up a whole new dimension of human-computer interaction.

For example, the iPhone is ideally suited for developing innovative health monitoring and support tools due to its mobility. This would benefit busy people who are always on the run as well as home-bound patients, such as many elderlies. Besides accelerometers, many mobile devices also support natural input through touch screens. These components allow us to design user interfaces for people that can be operated using their fingers, or rotating and moving the device.

Recently, we have developed and tested two iPhone games to support patients to perform rehabilitation exercises involving the arm's joints and muscles. The game fulfils two purposes at once: therapy and monitoring (Sunwoo et al., 2010). The games use the accelerometer, which allows users to control the game by swinging or flicking their device, similar to those on the Wii.

In one of the games, the player is presented with an environment similar to a real blowing game (see Figure 3). The game rules are based on the real bowling game, so that elderly people can enjoy playing a familiar game. The game is simply played by swinging the phone using the shoulder, just like what people would do in a real bowling game, hence exercising the shoulder joint.
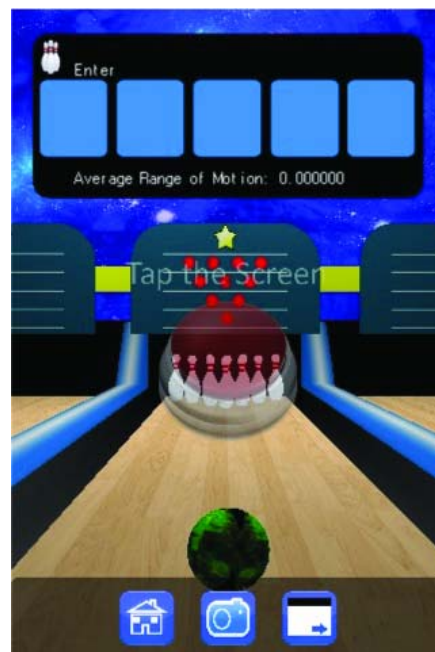
**Figure 3: Screenshot of the Bowling game
(Sunwoo et al., 2010)**

Wii remotes are inexpensive wireless input device. The Wii has a controller which has accelerometers inside and this allows actions to be converted into movement on the screen. Although it is primarily designed as an input device for the Nintendo Wii console, various other systems have used them as an input device. For example, Alankus et al. (2010) demonstrated the use Wii remote to detect arm motions for stroke rehabilitation.

Also standard webcam technology can be leveraged to support patients to perform rehabilitation activities at their homes. A low-cost webcam can be used to capture video data of a user's movement. Burke et al. (2009) reports the use of webcam games for stroke rehabilitation.   The games were designed to promote gross arm movements in order to aid upper limb rehabilitation. The main advantage of using the webcam is that the user is not required to have any experience with computers. This is advantageous for elderly as many of them face anxiety when confronted with technology.

6   Conclusion

We have analysed popular Web 2.0 health systems, classified them according to their main functionality and discussed some of their strengths and weakness. Existing patient-focused health systems are mostly based on information sharing, but provide only limited user interaction. They do not yet enable proactive healthcare for elderly users.

We proposed Healthcare4Life, which has a focus on enabling elderlies to manage their own health, by performing and monitoring rehabilitation activities. In this paper, we discussed how existing Web 2.0 resources and consumer devices can be leveraged to achieve this goal. As it turns out, there are technologies that allow us to integrate many of the functionalities of existing Health 2.0 systems. By adapting them to the special requirements of elderly people, Health 2.0 can be made more accessible to this important target group.

7   References

Anderson, P. (2007): What is Web 2.0? Ideas, Technologies and Implications for Education. JISC Technology & Standards Watch. http://www.jisc.ac.uk/media/documents/techwatch/tsw0701b.pdf Accessed 27 August 2010

Alankus, G., Lazar, A., May, M., Kelleher, C. (2010): Towards Customizable Games for Stroke Rehabilitation. *Proc of the 28th international conference on Human factors in computing systems*, Atlanta, Georgia, USA, 2113-2122, ACM.

Axon, S. (2010): Facebook's Open Graph Personalizes the Web, Mashable. http://mashable.com/2010/04/21/facebook-open-graph/ Accessed 27 August 2010.

Burke, J., McNeill, M., Charles, D., Morrow, P., Crosbie, J., and McDonough, S. (2009): Serious Games for Upper Limb Rehabilitation Following Stroke. VSGAMES '09, 103-110.

CureTogether: The Collective Wall, CureTogether – CureTogether.com, Inc. http://www.CureTogether.com/. Accessed 27 August 2010.

Carmichael, A. (2009): Crowdsourced Women's Health Books Released by CureTogether. Lifeboat Foundation. http://lifeboat.com/blog/tag/CureTogether Accessed 27 August 2010.

DailyBurn: Burn Fat, Lose Weigth and Gain Mussle – DailyBurn Daily Burn, Inc. https://dailyburn.com/. Accessed 30 August 2010.

DailyStrength: Online Support Groups and Forums at DailyStrength, DailyStrength, Inc. http://www.dailystrength.org/. Accessed 27 August 2010.

Domingo, M.C. (2010): Managing Healthcare through Social Networks. Computer **43**(7): 20-25.

Dishman E. (2004): Inventing wellness systems for aging in place. Computer, **37**(5):34–4.

Disaboom: Connecting the Millions Touched by Disability - Disaboom, Disaboom.com. http://www.disaboom.com/. Accessed 27 August 2010.

Elder Brief (2010): Emota: Bridging the Gap Between Emotion and Technology. Elder Brief. http://elderbrief.wordpress.com/2010/05/02/emota-bridging-the-gap-between-emotion-and-technology/ Accessed 27 August 2010.

Elkin, N. (2008): How America Searches: Health and Wellnes. iCrossing. http://www.icrossing.com/research/how-amer¬ica-searches-health-and-wellness.php.

Fox, S. (2006): Online Health Search 2006. Pew Internet & American Life Project. http://www.pewinternet.org/~/media//Files/Reports/2006/PIP_Online_Health_2006.pdf Accessed 27 August 2010.

Fox, S. (2009): The Social Life of Health Information. Pew Internet & American Life Project. http://www.pewinternet.org/Reports/2009/8-The-Social-Life-of-Health-Information/04-The-Social-Life-of-Health-Information/2-Social-networking-sites-are-used-only-sparingly-for-health-queries.aspx?r=1 Accessed 27 August 2010.

Jaffe, A. (2010): Facebook Social Plugins and Open Graph: What they mean for online newspapers http://www.editorsweblog.org/analysis/2010/04/facebook_social_plugins_and_open_graph_w.php Accessed 27 August 2010.

Jones, S., Fox, S. (2009): Generations Online in 2009. Pew Internet & American Life Project. http://pewresearch.org/pubs/1093/generations-online Accessed 27 August 2010.

Ko, M. N., Cheek, G. P., Shehab, M. And Sandhu, R. (2010): Social-Networks Connect Services. Computer **43**(8): 37-43.

O'Reilly, T. (2005): What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software. http://oreilly.com/web2/archive/what-is-web-20.html?page=1 Accessed 27 August 2010.

Onor, M. L., Trevisiol, M., Urciuoli, O., Misan, S., Bertossi, F., Tirone, G., Aguglia, E., Pascolo-Fabrici, E. (2008): Effectiveness of Telecare in Elderly Populations – A Comparison of Three Settings, Telemedicine and e-Health, **14**(2): 164-169.

PatientsLikeMe: Patients Helping Patients Live Better Every Day, PatientsLikeMe Inc. http://www.PatientsLikeMe.com/. Accessed 27 August 2010.

Sarasohn-Kahn, J. (2008): The Wisdom of Patients: Healthcare Meets Online Social Media. California Healthcare Foundation. http://www.chcf.org/~/media/Files/PDF/H/HealthCareSocialMedia.pdf Accessed 27 August 2010.

Singh, J., Wünsche, B. C. and Lutteroth, C. (2010a): Framework for Healthcare4Life - A Ubiquitous Patient-Centric Telehealth System. *ACM SIGCHI New Zealand Chapter's International Conference on Computer-Human Interaction*.

Singh, J., Wünsche, B. C. and Lutteroth, C. (2010b): A Taxonomy of Usability Requirements for Home Telehealth Systems. *ACM SIGCHI New Zealand Chapter's International Conference on Computer-Human Interaction*.

Smith, J. (2010): December Data on Facebook's US Growth by Age and Gender: Beyond 100 Million. Inside Facebook. http://www.insidefacebook.com/2010/01/04/december-data-on-facebook's-us-growth-by-age-and-gender-beyond-100-million/ Accessed 27 August 2010.

Smith, J. (2009): Facebook Crosses 60 Million Monthly US Users, But Fewer People Over 55 Coming Back. Inside Facebook. http://www.insidefacebook.com/2009/05/26/facebook-crosses-60-million-monthly-us-users-but-fewer-people-over-55-coming-back/ Accessed 27 August 2010.

Sunwoo, J., Yuen, W., Lutteroth, C. and Wünsche, B. (2010): Mobile Games for Elderly Healthcare. ACM

SIGCHI New Zealand Chapter's International Conference on Computer-Human Interaction.

Swan, M. (2009): Emerging Patient-Driven Health Care Models: An Examination of Health Social Networks, Consumer Personalized Medicine and Quantified Self-Tracking, Int. J. Environ. Res. Public Health, **6**: 492-525.

Tullis, T. and Albert, B. (2008): *Measuring the User Experience - Collecting, Analyzing, and Presenting Usability Metrics*. Elsevier, Morgan Kaufmann.

Wauters, R. (2010): IAC Buys Into Fitness Social Network DailyBurn. TechCrunch. http://techcrunch.com/2010/05/20/iac-dailyburn/ Accessed 27 August 2010.

Weber, G. (2008): A Platform-Independent Approach for Auditing Information Systems. Second Australasian Workshop on Health Data and Knowledge Management (HDKM 2008). Wollongong, NSW, Australia. **80**: 65-73.

Weitzel, M., Smith, A., Lee, D., Deugd, S., Yates, R. (2010): A Web 2.0 Model for Patient-Centered Health Informatics Applications. Computer **43**(7): 43-50.

Weitzel, M., Smith, A., Lee, D., Deugd, S., Helal, S. (2009): Participatory Medicine: Leveraging Social Networks in Telehealth Solutions. In: Mokhtari, M., Khalil, I., Bauchet, J., Zhang, D., Nugent, C. (eds.) ICOST 2009. LNCS, **5597**: 40–47. Springer, Heidelberg.

Real Time Processing of Data from Patient Biodevices

Rhodora Abadia[1], Andrew Stranieri[2], Anthony Quinn[2], Sattar Seifollahi[2]
[1]William Light Institute, South Australia
[2]Centre for Informatics and Applied Optimisation
University of Ballarat
Ballarat , Victoria 3353

`rabadia@wli.sa.edu.au, a.stranieri@ballarat.edu.au`

Abstract

Patient biodevices worn by the infirmed detect vital signs and can help to improve health outcomes and the efficient provision of care. The streaming data generated by these devices can result in extremely large flows which cannot be analysed with existing data mining approaches. This paper surveys stream mining research and advances a new approach for the analysis of real time data generated by patient monitoring devices. The approach is computationally simple to scale up to process streams comprising huge volumes of data.

Keywords: stream data mining, sub-sequence counting, patient monitoring

1   Introduction

The ageing population is one of the major challenges being experienced by Australia's population. Health expenditure is greatest for the elderly and this leaves a great need for a health and aged care system (Australian Institute of Health and Welfare 2010). Patient biodevices worn by the infirmed or elderly to detect vital signs help improve health outcomes and efficient provision of care. These enable early detection of high risk states and can trigger alerts. These devices function by continuously streaming data to computer programs that analyse a historical sequence of observations and relate the data to characteristics of the patient, recent treatments and conditions diagnosed.

The streaming data generated by these devices result in extremely large flows that are typically usually too huge to be stored on permanent devices. The data rate of these streams is not constant and the data stream patterns are continuously evolving. Existing data mining approaches are not sufficient to analyse and detect significant patterns or intelligently raise alarms regarding the patient's condition. A different approach for examining these kinds of data being passed by these patient devices is needed.

Stream data mining is the task of discovering interesting and hidden patterns from a continuous, rapid and often large volume of data. The challenge is to adapt queries and perform mining tasks over rapid data streams. The application of stream data mining in the analysis of data from large number of patients generated by biodevices presents an opportunity to lead to insights into many conditions, accidents and precursors to adverse events.

In this paper, we introduce an approach for analysing streamed data that is computationally simple to scale up to be able to process streams comprising huge volumes of records. This method promises to facilitate some insight to patterns that should trigger alarms. This paper proposes an extension of the AwSum algorithm (Quinn et al 2008). AwSum uses conditional probabilities that demonstrate that the concept of weighted sum provides a simple and intuitive way to understand how feature values influence each other. In the approach advanced here, AwSum is applied to streaming data from patient monitoring devices. Our approach is based on sketching and involves the intuition that the frequency of occurrence of all sub-sequences that exist in a sequence can be obtained by incrementing the counts of each sub-sequence. A run time construction of the tree is created as the data is streamed in.

This paper is organised as follows. In the next section, we will discuss the different approaches to data summarisation and streaming data. Section 3 briefly surveys on recent mining patient monitoring data. Section 4 discusses the proposed approach for the analysis of real time data. In section 5, variations of the proposed approach are presented. The conclusion and limitations of the approach are discussed in section 6.

2   Approaches to Stream Data Mining

Data streams pose unique space and time constraints on the computation process (Aggarwal and Yu 2007). Unlike conventional data mining, the volume of data arriving in streams is too large to store in databases. Consequently, stream mining approaches must deal with problems associated with limited memory space. A further consequence is that the processing of the mining algorithm must occur in real time which challenges computational processing efficiency. This section reviews data summarisation methods that address the storage challenge and stream mining algorithms that address the computational challenges.

2.1 Data Summarisation

In order to answer the challenges of resource constraints in stream data mining, many data summarisation techniques have been adapted from statistics. In the application of data streams, sampling methods are still considered among the simplest methods for identifying data that can be used for analysis. Sampling is a statistical method which, when applied to data processing, refers to the process of identifying the probability of a data item being processed or not. Sampling methods are effective for multi-dimensional data sets. However, using sampling in the analysis of streaming data poses a problem of unknown data sets.

Variations of sampling methods such as reservoir-based sampling methods by Vitter (1985) and concise sampling advanced by Aggarwal and Yu (2007) addressed this problem. The Reservoir-based method was developed to do a one-pass access of data from magnetic devices and was recently adapted to streaming data. The idea is to pick out a sample (reservoir) from the stream with a pre-decided size and apply stream mining algorithms to estimate results. Since data streams evolve over time, query results may also change over time. The reservoir window was recently adapted to get data from a moving window (Babcock, Datar, Motwani, O'Callaghan 2002). However, this approach is useful for data stream applications where a small amount of recent history is more relevant than the entire data stream. The concise sampling method (Aggarwal and Yu 2007) is another extension of the reservoir window method. This method improves the effectiveness of the reservoir window and is most applicable for one variable sampling in a single dimension.

Even though variations of sampling methods address the problem of unknown data set size, these methods do not address the problem of fluctuating data rates. Relationships between sampling rate, data rate and error bounds need to be investigated (Gaber, Zaslavsky, and Krishnaswamy 2005).

The Wavelet is a hierarchical data decomposition and summarisation technique whose theoretical foundation relies on the fact that data objects can often be represented as piecewise linear functions (Keim and Heczko 2001). Signal processing, data compression and computer graphics are the most common applications of wavelets. Wavelet decomposition has been used with streamed data in large data networks by Bulut and Singh (2003) where a central site summarises while the client sites ask queries. The wavelet decomposition is also used with problems of multiple measures associated with time series. An example of a multiple measure problem is when a patient's biodevice is used to simultaneously track the patient's blood pressure, oxygen level, and temperature at a given instance of time. However, the simple wavelet approach of decomposing along different measures and selecting the largest co-efficient for each measure has proven to be inefficient. A proposed approach is to use an extended wavelet representation to simultaneously track multi-measure coefficients of the wavelet representation rather than selecting only one (Degligiannakis, Garofalakis and Roussopoulos 2007).

Sketching is another approach to data summarisation. It is an extension of the random projection technique which identifies a subset of the features of a given data stream. It is a probabilistic counting method which uses a randomised function to perform counting on the data stream. This technique is often used in determining representative trends in the time series domain (Indyk, Koudas and Muthukrishnan 2000). Sketching has been applied to problems of identifying most frequent items in the data streams (Charikar, Chen, Farach-Colton 2002) and finding significant differences over data streams (Fong and Strauss 2000). One advantage of sketching is that it has extremely low space requirements. The problem with sketching is that most of the methods are based on the analysis of a single dimensional stream of data points which makes it difficult to apply to most data streaming problems because they are naturally multi-dimensional. Further, sketching is good for determining large aggregate signals but is not useful for counting infrequently occurring items in the data stream (Aggarwal and Yu 2007).

The histogram is also a data summarisation technique. It divides the data along any attribute into a set of ranges and maintains the count for each set (i.e., buckets). One problem with the use of histograms is that the distribution of data points within a bucket is not retained and therefore assumed to be uniform (Ioannidis 2003). This is a key source of inaccuracy when applied to data streams.

2.2 Stream Data Mining Techniques

Over the past few years, several algorithms for extracting knowledge from data streams have been developed. These techniques were already being used in data analysis but were modified to analyse streaming data. It is worth mentioning that many stream data mining techniques use conditional probabilities in their approach to streaming data analysis. This section discusses the most commonly used stream data mining techniques: clustering, frequent pattern, and classification.

Clustering is a common statistical data analysis approach that is used in data mining and machine learning. This approach finds a partition of data points so that similar items are in the same part of the partition while different items are in a different part. Different variations of the *k-median* clustering technique have been proposed in streaming data analysis. Most of the work focused on providing a solution to the *k-median* and *k-center* clustering problems approach in the data stream context. The *k-center*'s objective is to minimize the largest assignment distance while *k-median*'s objective is to minimise the sum of assignment distances. Guha et al (2003) proposed a one pass clustering of data streams algorithm which uses a divide and conquer approach so that clustering can be done in a small space. The problem of increasing approximation and the problem of merging clusters that are far apart are the main limitations of this approach. A variation of the k-median algorithm by Charikar, O'Callaghan and Panigrahy (2003) that addresses the limitation of an increasing number of levels used in the divide and conquer approach has been proposed. Exponential histogram (Babcock, Datar, Motwani and O'Callaghan 2003) was used to answer the

limitation of merging clusters that are far apart. Other clustering approaches include micro-clustering advanced by Aggarwal et al (2003), the use of *k-motifs* to choose the subsequences to produce meaningful results (Keogh, Lins, and Truppel 2003) and lightweight clustering (Gaber, Krishnaswamy, Zaslavsky 2005).

Most of the frequent pattern mining has high memory requirements when the number of distinct items is large and/or the support level desired is quite low. In data streams, some proposed methods to answer this problem find frequent item sets over a sliding window or the entire data stream (Gianella, Han, Pei, Yan, and Yu 2002; Jin and Agrawal 2005). In sliding window approaches analysis is performed on the most recent data stream and summarized versions of the old data items. (Gianella, Han, Pei, Yan, and Yu 2003) uses the sliding window in their proposed frequent pattern approach. A modified frequent pattern tree (FP-stream) is built incrementally to summarise the frequent patterns at multiple time granularities. Jin and Aggrawal (2005) use a single pass algorithm. The proposed algorithm uses a prefix tree using a hash table to provide an efficient use of memory. Another approach to frequent pattern mining uses approximate frequency counts that use historical data to incrementally count frequency pattern (Manku and Mowani 2002).

Classification is a classic data analysis approach which has roots from machine learning. It is used to predict group membership (or classes) for data instances. Classification in the context of data stream mining is made more difficult by the constant changes in the underlying data stream. Classification algorithms typically evaluate their performances on data streams with balanced class distributions. A general framework using classifications in mining concept drifting data streams have been proposed using weighted classifiers (Wang, Fan, Yu, Han 2003). This approach assumes that the "concept drift as a probability of the change in a class given a feature (ρ (class/feature)). In reality, changes can only be observed as a joint probability and it is difficult to identify whether the changes are caused by the changes in the feature or the joint probability. A proposed improvement to this approach is the classification of data streams with skewed class distribution (Gao, Fan, Han, Yu 2007). This approach assumes that mining data chunks improves data accuracy over time; descriptive model is preferred over generative model for general purpose stream mining; and that estimating probability is more suitable than label predictions. Several decision tree classifiers have been proposed either as a single tree or a collection of such trees. CVFDT (Hulten Spence and Domingos, 2001) is a known algorithm for building a single incremental decision tree. It uses the Hoeffding bounds to decide whether another level of the tree needs to be created. With this type of approach, accuracy is dependent on a large number of samples and it takes time to update the tree. Having small number of samples result in unsatisfactory performance. Using collections of decision trees (Wang, Fan, Yu, and Han 2003, Street and Kim 2001) provide better performance than single classifiers. However, ensemble methods in decisions trees often provides more complex models (Han and Kamber

2006) which makes it difficult to understand the problem at hand.

3 Mining patient monitoring data

Alemdar and Ersoy (2010) survey sensor networks in health care and identify categories of applications including monitoring falls, identifying daily activities, location tracking, medication tracking and medical status monitoring. Hardware challenges include making devices unobtrusive, issues with their energy consumption and reliable transmission of data to servers. Application level challenges include issues related to security and computational efficiency of programs that process the data streaming in. Ramon et al (2010) demonstrate that data mining can enhance the processing of data from intensive care patients though identify significant challenges. One difficulty described above involves applying mining algorithms to enormous datasets. Another difficulty involves learning thresholds from datasets of readings from many patients where many thresholds are unique to each patient. For example, a blood pressure of 160/90mmHg may be alarming for one patient yet normal for another. Lee et al (2010) present an example of the potential data mining can bring to monitoring by integrating environmental data including weather information with patient bio-data to predict an asthma attack. In that study, Bayesian classifiers are used to learn sequences of patterns that are significant indicators or an asthma attack for a patient. In that approach, the mining is performed off line because of the computational complexity inherent in calculating and revising the necessary conditional probabilities as the data arrives at each moment in time. In the next section, we present an approach that aims to calculate the counts of each sub-sequence present in a sequence, in real time. Conditional probabilities can be readily calculated from the sub-sequence counts.

4 Sub-sequence counting

The stream mining approach advanced here does not store any data permanently. Instead, counts of sub-sequences are kept and, once obtained, can be used in any mining algorithm that deploys conditional probabilities including association rules and Bayesian classifiers. The challenge is to arrive at an algorithm that can increment a count of all or a sufficient number of sub-sequences in real time. Our approach involves the intuition that the frequency of occurrence of all sub-sequences that exist in a sequence can be obtained by incrementing the counts of each sub-sequence of length 1, 2, …,d to a user set threshold length, d. Once the sub-sequence counts are obtained they can be converted to conditional probabilities for use in probabilistic classification algorithms. For instance, in the forecasting algorithm described by Bertoli and Stranieri (2004), counts for sub-sequences *abbc, bbc, bc* from a stream lead to the conditional probabilities pr(*c/abb*), pr(c|bb) and pr(c|b). These are combined to generate forecasts.

Sub-sequence counts are converted to conditional probabilities for the determination of an influence score in the classification and visualisation algorithm advanced by Quinn et al (2008). In this supervised learning algorithm, the pr(class value$_1$ | feature value$_1$) and pr(class value$_2$ |

feature value$_1$) are used to calculate a score that represents the influence of a feature value on a class variable. For example, the influence score for the feature gender = female on a cystic fibrosis severity data set used by Quinn et al (2008) indicates this feature plays a small influence toward mild forms of the disease. The influence score of the presence of a yeast infection, candida albicans indicates this has a small influence toward severe cystic fibrosis. However, taking the pair together, female and candida albicans, has a large influence toward severe cystic fibrosis.

That algorithm generated the conditional probabilities of all single feature values, pairs and triples and demonstrated comparable accuracy with other supervised learning approaches. The approach involved a single pass of the data to generate sub-sequence counts but assumed all data was stored in a database. The approach described here involves the generation of sub-sequence counts as the data streams in, one item at a time.

This is done with the run time construction of a tree where the first level of the tree stores the counts for sub-sequences of length 1, the second stores the values of sub-sequences of length 2, of a features values, up to the dth level that stores the longest sub-sequences that we wish to record counts for. The latest data item to arrive from a stream adds a new sub-sequence of each length which is inserted into the tree at each level. The counts of each sub-sequence are ultimately read by traversing the tree.

The algorithm is illustrated here by way of a simple example. Imagine a very short stream of 5 real valued oxygen concentration data from t1 to t5: 96%, 96%, 94%, 88%,88%. A tree is initialised where the first level in the tree will be used to store the counts of sub-sequences of length 1, the second level will be used to store the counts of sub-sequences of length 2 and so on. The depth of the tree is limited by a user set threshold.

The first item, 96 is inserted into the tree as a new arc representing 96. The node at the end of the arc represents the frequency that 96 has occurred to date. The node is tagged with a * as the node that was most recently updated at that level as illustrated in Figure 1.
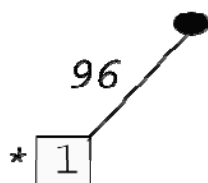

Figure 1 First reading inserted

The next value to be read in is 96%. In this case, the count for the length 1 sequence, 96, is incremented and a new length 2 sequence; 96, 96 is captured by the addition of a new arc from the node at the end of the 96 arc as illustrated in Figure 2.
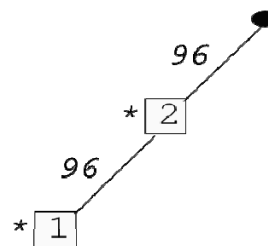

Figure 2 Tree after the first two entries inserted

The next item, 94 is now read. An arc for 94 does not exist at level 1, or 2 so is created and counts are incremented accordingly as illustrated in Figure 3.
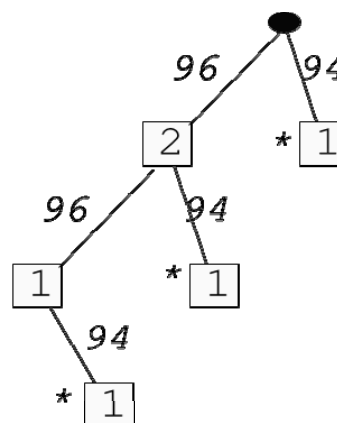

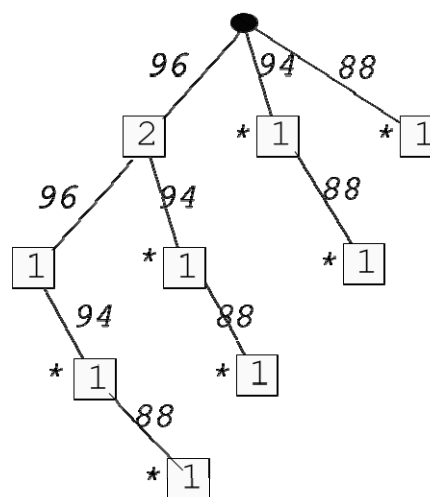Figure 3 Tree after the first three entries inserted

Figure 4 Tree after the first four entries inserted

The frequency of all subsequences in the sequence to date is generated from a traversal of the tree to yield the counts illustrated in Table 1.

The number of arcs from the root node equals the number of distinct single characters read in. The number of arcs from level 1 nodes equals the distinct number of pairs and the number of arcs from level 2 nodes equals the number of distinct triples.

The size of the tree depends on the number of unique sub-sequences. A sequence with no repeating entries (i.e. a

series) represents the worst case and a sequence with a single reading represents the best case. A worst case sequence of length n, will result in n nodes at level 1, and n-1 nodes at level 2 and n-2 nodes at level 3 and so on.

Sequence	Length	Count
96	1	2
94	1	1
88	1	2
96,96	2	1
96,94	2	1
94,88	2	1
88,88	2	1
96,94,88	3	1
94,88,88	3	1
96,94,88,88	4	1

Table 1 Subsequence counts

For a worst case tree, each new entry requires the creation of a new single node at each level. Although each creation is plausibly performed in parallel, the number of operations increases with the length of the sequence. The size, S, of the worst case tree is:

$$S = \sum_{1}^{d} n_i = \sum_{1}^{d} (n - i + 1)$$

where:
d is the depth of the tree
n_i is the number of nodes at level i.

The best case tree derives from a sequence length n with a single character repeating n times. The best case tree has a single node at each level and has n levels.

The average case is empirically determined and can be expected to vary from dataset to dataset.

The size of the tree for the average case is associated with the proportion of repeating entries which will vary from dataset to dataset.

5   Variations and enhancements

The number of occurrences of a sub-sequence can in general be expected to diminish with the length of a sub-sequence. In a stream from a patient bio-device, there are as many sub-sequences of length 1 as there are distinct observations. So, for the sample stream above, there are 2 occurrences of 96, and 1 each of 94 and 88. There are fewer sub-sequences of length 2. In a diabetes dataset described by Quinn et al (2008) with 78 features and 1600 records, 94% of the pairs of items possible were actually observed in the dataset and 74% of all combinations of three items possible were observed in the data. This suggests that long sub-sequences do not repeat often in

practice unless the data has long cycles. A consequence of this is that the number of unique sub-sequences will not, in general be high so the algorithm presented here will only increment counts on existing sub-sequences rather than enlarge the tree with new arcs. This suggests that the memory requirements of the tree will be manageable even with long streams of data.

6   Conclusion

Monitoring of patients with the use of wireless bio-devices is an emerging trend as populations age and cost effective treatment requires increased home based care. There are many challenges inherent in establishing wireless patient monitoring, including difficulties associated with the processing of data as it arrives in real time. Data mining shows a great deal of promise in this, yet conventional approaches assume more memory and processing resources than is typically available. Approaches that are based on conditional probabilities including association rules and Bayesian classifiers can require less memory and processing resources than other approaches yet calculating and revising the conditional probabilities in real time as data items arrive is difficult. In this paper, endeavours in stream mining are surveyed and a new approach is advanced that enables conditional probabilities to be revised as data items arrive. Future work aims to evaluate the extent to which the approach advanced is computationally efficient in real world contexts.

7   References

Alemdar, H., Ersoy, C., Wireless sensor networks for healthcare: A survey, Comput. Networks. (2010) 10.1016/j.comnet.2010.05.003

Australian Institute of Health and Welfare (2010). Health System Expenditure on Disease and Injury in Australia, 2004–05. AIHW Cat. No. HSE 87. AIHW Health and Welfare Expenditure Series no. 19.   Canberra, Australia.

Aggarwal, C., Han, J. ,Wang, J., Yu, P. (2003). A Framework for Clustering Evolving Data Streams. *In Proceedings of the 29th International Conference on Very Large Databases.* 29: 81-92. VLDB Endowment.

Aggarwal, c. Han,J, Wang , J. Yu, P (2004). On Demand Classification of Data Streams. *In Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 503-508. ACM, New York.

Aggarwal, C. and Yu, P. (2007): A Survey of Synopsis Construction in Data Streams. In   *Data Streams: Models and Algorithms.* 169-207. Aggarwal, C. (ed.). Springer, London.

Babcock, B., Datar M., Motwani, R. and O'Callaghan, L. (2003). Maintaining Variance and k-Medians over Data Stream Windows, *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems.* 234-243. ACM, New York.

Bertoli, M and Stranieri, A., (2004). Forecasting on complex datasets with association rules. *8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems. Wellington.* 1170-80

Bulut, A., Singh, A. (2003). Hierarchical Stream Summarization in Large Networks. *Proceedings of the 19th International Conference on Data Engineering.* 303-314. IEEE Computer Society.

Charikar, M., Chen, K., Farach-Colton, M. (2002). Finding Frequent Items in Data Streams. *Lecture Notes in Computer Science. Proceedings of the 29th International Colloquium on Automata, Languages and Programming.* (2380) 693-703.

Charikar, M., O'Callaghan, L., Panigrahy, R. (2003). Better Streaming Algorithms for Clustering Problems. *In Proceedings of the 35th Annual ACM Symposium on Theory of Computing.* 30-39. ACM, New York.

C.-H. Lee, et al.,(2010) A novel data mining mechanism considering bio-signal and environmental data with

applications on asthma monitoring, Comput. Methods Programs Biomed.

doi:10.1016/j.cmpb.2010.04.016

Degligiannakis, A., Garofalakis, M. and Roussopoulos, N. (2007). Extended Wavelets for Multiple Measures. *ACM Transactions on Database Systems.* 32 (2) Article No. 10. . ACM, New York.

Fong, J., Strauss, M. (2000). An Approximate L_p-difference Algorithm for Massive Data Streams. *Lecture Notes in Computer Science. Proceedings of the 17th Annual Symposium on Theoretical Aspects of Computer Science.* (1770) 193-204. Springer-Verlag.

Gaber, M.M. , Zaslavsky, A. and Krishnaswamy, S. (2005). Mining Data Streams: A Review. *ACM SIGMOD Record.* 34(2): 18-26. ACM, New York.

Gao, J., Fan, W., Han, J. and Yu, P (2007). A General Framework for Mining Concept-Drifting Streams with Skewed Distribution. *SIAM International Conference on Data Mining (SDM'07).*

Gianella, C. Han, J., Pei, J. Yan,X., and Yu, P (2003). Mining Frequent Patterns in Data Streams at Multiple Time Granularities. *Next Generation Data Mining.*Kargupta, H., Joshi, A. .Sivakumar, K. and Yesha, Y/ (eds.) 191-212. AAAI/MIT.

Guha, S., Meyerson, A., Mishra, N., Motwani, R. and O'Callaghan, L. (2003) Clustering Data Streams: Theory and Practice. *The Knowledge Discovery (TKDE) Special Issue on Clustering.* (15).

Han, J and Kamber, M. Data Mining – Concepts and Techniques- 2nd Edition (2006). Morgan Kaufmann Publishers. 291-306

Hulten, G., Spencer, L., Domingos, P. (2001) Mining Time Changing Data Streams. *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 97-106. ACM. New York.

Indyk, P., Koudas, N., and Muthukrishnan, S. (2000). Identifying Representative Trends in Massive time Series Data Sets Using Sketches. *Proceedings of the 26th International Conference on Very Large Data Bases.* 363- 372. VLDB Endowment.

Ioannidis, Y. The History of Histograms. *Proceedings of the 29th International Conference on Very Large Databases.* 29: 19-30. VLDB Endowment.

Jin, R., Agrawal, G. (2005). An Algorithm for In-Core Frequent Itemset Mining on Streaming Data. *In Proceedings of the Fifth IEEE International Conference on Data Mining.* 210 - 217 . IEEE Computer Society.

Keim, D.A., Heczko, M. (2001). Wavelets and their Applications in Database. 17th *International Conference on Data Engineering.* Heidelberg, Germany.

Keogh, E., Lin, J., Truppel, W. (2003). Clustering of Time Series Subsequences in Meaningless: Implications for Past and Future Research. *In Proceedings of the 3rd IEEE International Conference on Data Mining.* (8) 154-177. IEEE Computer Society.

Manku, G.S., and Motwani, R. (2002) Approximate frequency counts over data streams. *In Procoddings of the 28th International Conference on Very Large Databases.* 346-357. VLDB Endowment

Quinn, A: Stranieri, A., Yearwood, J, Hafen G. and Jelinek, H. (2008) AWSum-Combining Classification with Knowledge Acquisition. International Journal of Software and Informatics. 2(2) 199-214.

Street, W., and Kim, Y. (2001). A streaming ensemble algorithm (SEA) for Large-Scale Classification. *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 377-382. ACM. New York.

Tseng, Vincent, S., Lee-Cheng Chen, Chao-Hui Lee, Jin-Shang Wu, Yu-Chia Hsu, "Development of a Vital Sign Data Mining System for Chronic Patient Monitoring," cisis, pp.649-654, 2008 International Conference on Complex, Intelligent and Software Intensive Systems, 2008

Vitter, J.S. (1985) Random Sampling with a Reservoir. *ACM Transactions on Mathematical Software,* 11(1): 37-57.ACM, New York.

Wang,, H, Fan, W., Yu, P, and Han, J. (2003). Mining Concept-Drifitng Data Strams Using Ensemble Classifiers. *Proceedings of the 9th ACM Internationalconference on Knowledge Discovery and Data Mining.* 226-235. ACM, New York.

Coupling of Indigenous-patient-friendly cultural communications with Clinical Care Guidelines for Type 2 Diabetes Mellitus

David Forbes, Amandeep Sidhu and Jaipal Singh

Digital Ecosystem and Business Intelligence Institute

Curtin University

Perth, Australia

`david.e.forbes@postgrad.curtin.edu.au, {a.sidhu, j.singh}@curtin.edu.au`

Abstract

Distance, terrain, climate and inadequate medical resources seriously constrain health care accessibility for rural and remote Indigenous communities of Western Australia (WA). Management of the Type 2 Diabetes Mellitus (T2DM), a chronic condition affecting Indigenous people much more than non-Indigenous, requires a complex assortment of time-sensitive communications activity and interventions to avert serious complications. Communications barriers arising from pervasive cultural misunderstanding in primary care go far beyond language differences and routine translation techniques. Practitioners and patients lacking the ability to share understanding in the examination and testing discourse need a culturally sensitive purpose-driven informatics system of support for the Patient-Practitioner Interview Encounter (PPIE). The dominant unidirectional clinician-biased forms of communication employed by healthcare professionals are a major barrier. Our developing communications support model utilizes the mapping of ontologies. The Community Healthcare ontology is dedicated to mapping a clinical taxonomy for T2DM national guidelines to Aboriginal English (AE). The eventual user interface will represent Aboriginal patient-culture use of interactive audio visual media in the primary healthcare setting. This research objective establishes respect for the Aboriginal patient's dialectal and pragmatic preferences, enabling us to couple these preferences with Australia's Standard English clinical communications practice for treatment and care of Indigenous T2DM patients. A critical capability of the eventual application, especially when phrase ontology guidance enters the interface will be the interception of ambiguities and mitigation of misinterpretation risk. The emphasis is concentrated on bi-directional communications assistance that will not only enhance the Aboriginal patient opportunity to contribute to the PPIE, but will reinforce the value of and reciprocal respect for, sound clinical practice.

Keywords: Indigenous Health, Health Communication, Aboriginal English Ontology, Patient-Practitioner Interview Encounter, Type 2 Diabetes

1 Introduction

The complexities, volume and time consumption factors of primary care patient evaluation processes are presenting healthcare practitioners globally with extraordinary decision-making demands (Barlow, 2009, Fiscella and Epstein, 2008). Access to healthcare information via the Web is ubiquitous for many people in developed countries; and interpretation by non-medically-trained patients can potentially help or hinder patient-practitioner interactions (Seckin, 2010). We suggest that this is further contributing to healthcare inequities affecting patients who are socio-culturally disadvantaged, as advancements in healthcare literacy are increasingly influenced by self-education driven by access to and affordability of, Information Communications Technologies (ICT). In the primary healthcare setting, complexity includes volume, time-sensitive analysis and choice of treatment and care options, heavily influenced by clinician service-time constraints. This presents a strong case for actively seeking cost efficient smart two-way interview support systems, introducing ICT models that will enrich the disadvantaged patient communications capabilities while alleviating the risk of clinician task overload and burn out. In this work we are focused on patient-provider interaction involving Western Australia Aboriginal Type 2 diabetes patients, particularly those who are additionally disadvantaged by the healthcare service delivery constraints of rural and remote areas within this large, demographically unbalanced, logistically challenged state.

2 Background

WA's remote area Indigenous T2DM patients are disproportionately at risk of developing chronic diseases compared with non-Indigenous people living in or close to urban areas (ENDOCRINE HEALTH NETWORK, 2008). Western Australia has the largest land area (2,532,400sq km) of any Australian State or Territory. Its coastline of 12,500km amounts to 34% of Australia's total coastline. Over 72% of Western Australia's population is located in Perth, where principal health care support facilities, medical treatment and pathology testing services are located.

Diabetes is one of the most serious chronic diseases prevalent in the world today. The World Health Organization (WHO 2009) describes chronic diseases as 'diseases of long duration and generally slow progression'. Type 2 Diabetes Mellitus (T2DM) comprises 80% to 90% of all diabetes cases (WHO 2009). The first accurate diabetes prevalence study

commissioned in Australia led to a 2001 report by the International Diabetes Institute suggesting that almost one in four Australians aged 25 years and older had either diabetes or a condition of impaired glucose metabolism (International Diabetes Institute, 2001).

2009 statistical data (ABS 2010, HealthInfoNet 2010) show the high incidence of diabetes in the Indigenous population compared with non Indigenous.

(a) Proportions of people reporting diabetes/high sugar levels with Indigenous to non-Indigenous ratios, Australia, for the year 2004-2005, indicate that 1% of Indigenous people aged between 15-24 have diabetes, versus 0.5% of non-Indigenous population of the same age, a ratio of 2.

(b) In the age group 25-34 the (2004-05) percentages were 4.3% Indigenous, 0.6% non-Indigenous, a ratio of 7.2

(c) In the age group 55 years+ the percentage of the Indigenous population with diabetes (2004-05) was 32.1% compared with 11.2% of the non-Indigenous population, a ratio of 2.8

Diabetes WA reports that in the 2007/08 financial year, 12.37% of all hospitalisations in Western Australia, i.e. 95,775 admissions, involved patients with diabetes, whether as a direct or associated condition (DIABETES WA, 2009).

Rural and remote communities rely upon thinly-spread, mobility-dependent, over-stretched, ill-equipped and sometimes inaccessible, primary care resources; and on a relatively small cohort of Aboriginal Health Workers of Indigenous ethnicity. These adverse factors are compounded by comparably weak communications infrastructure, and sporadic development of telehealth services (M.Cribbs and K.Glaister, 2007, P.Van Ast and A.Larson, 2007)

For urban, rural and remote region T2DM patients, the common initial evaluation locus is the primary care environment. Convenient and timely access to a General Practitioner (GP) and referral to specialist support diminishes with remoteness, with extreme contrast in levels of service and capabilities experienced by remote patients compared with their urban counterparts. Within the innumerable coordinates of T2DM and co-morbidity management for Indigenous patients, primary care communications occupies a key quality status for reliable evaluation of care priorities. There is evidence to show that patient-practitioner interactions are falling short due to a cultural disconnect leading to misunderstandings between clinicians and their Aboriginal patients (Carrillo et al., 1999, Kagawa-Singer and Kassim-Lakha, 2003, Shahid et al., 2009, Lawrence et al., 2009).

We have concluded from this research that for a variety of reasons applicable to both Indigenous and non-Indigenous chronic disease patients, the dominant unidirectional clinician-biased forms of communication employed by healthcare professionals are a major barrier to communications and thereby contribute to lower efficacy in patient evaluation and care outcomes. Modelling for elevation of patient communication

capabilities, potentially a bi-directional and constructive patient empowerment augmentation of existing clinical care is an attractive route to counter this.

Specifically in the context of Aboriginal patient communication, there is an intricate montage of socio-cultural nuance that is not widely understood, and therefore not appropriately valued and respected, by a sizeable percentage of WA's healthcare providers. Language translation and the use of interpreters, while having some worth, are limited in both availability and value. Literal translation in isolation contains risk characteristics that may exacerbate misunderstanding (Trudgen, 2000). The healthcare sector also has its own specialist culture to add further distance to differences in participant ethnicity and socio-economic background. The cultural disconnect(s) can lead to inadequate information exchange, zero information exchange, potential but unqualifiable relevance/irrelevance in the exchange, misinformation, misinterpretation, and misunderstanding during patient-practitioner encounters (Trudgen, 2000). This emphasizes the need to recognize that technology must go beyond technical capability to promote and augment productive PPIE relationships. The communications deficit for Indigenous peoples globally is apparent in disease treatment and care protocols other than T2DM. Reporting on patients' views about effective communications Shahid et.al highlights treatment disparities affecting Aboriginal cancer patients in Western Australia hospital settings (Shahid et al., 2009). ICT is viewed by other researchers as a means to help overcome these disparities through communications enhancement. Automation prospects include the use of medical dialogue management systems (Beveridge and Fox, 2006); and multi-layered conversational intelligent agent systems (Goh et al., 2007). In the absence of such developments specific to Australian Aboriginal PPIE support, we have embarked upon the design of a system that establishes value of and respect for the Aboriginal patient's dialectal and pragmatic preferences, thereby enabling us to couple these preferences with Australia's Standard English clinical communications practice in the treatment and care of Indigenous T2DM patients.

3   Methodology

Our developing model utilizes the mapping of ontologies. A Community Healthcare ontology is dedicated to mapping a clinical taxonomy for T2DM national guidelines (DIABETES AUSTRALIA, 2009) to Aboriginal English (AE) ontology. The knowledge representation outcome will be achieved through formal semantics and syntax applications using the Web Ontology Language (OWL).

AE is a dialectal form of communication that with some individual word variation is used within Australian Aboriginal communities nationwide where Standard Australian English has become the dominant local language of society's establishment, i.e. schools, clinics, hospitals. Most commonly it represents the uniquely Aboriginal conversational modality that takes place in the family home, and is often cross-referred as 'home talk'. Harnessing AE in readiness for two-way mapping with Standard English T2DM guideline terminology through

ontology development requires validation by AE speakers from within Aboriginal communities. This process will be conducted and managed by Aboriginal focus groups. The eventual user interface is not discussed here, save to say that it will represent Aboriginal patient-culture-driven access to and use of interactive audio visual media to correlate and merge digitally generated assessment test data with information that reflects the bi-directional PPIE. Systems that induce a mutually cooperative PPIE are likely to favour Aboriginal-friendly environment designed touch screen facilities. There is evidence to show that Aboriginal communities willingly embrace and culturally adapt their usage of touchscreen technology to their ultimate advantage (Auld, 2002).

The AE ontology tree at Figure 1 illustrates the ontology scope as at this writing. This conveys the considerable range of AE communications structure and pragmatics, with some latitude for expansion still remaining. For example, instances of subclass properties under *Pragmatics* and also under *Conversational Modality* may include PPIE Phrases, divided into *AE Speaker Sent* and AE *Speaker Received* modes.
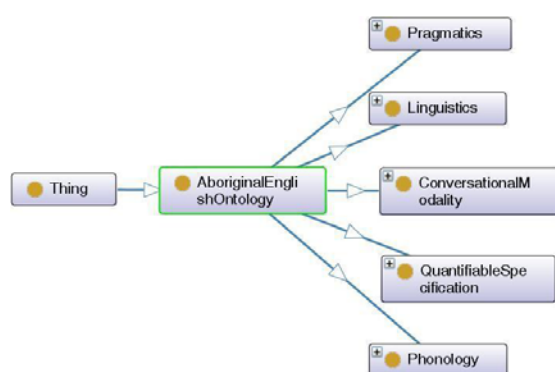

**Figure 1: Aboriginal English Ontology**

A critical capability of the eventual application, especially when phrase ontology guidance enters the interface will be the interception of ambiguities and mitigation of misinterpretation risk. The work discussed here dwells on this aspect. In AE the risk of ambiguity and misunderstanding is pervasive. For example research has shown that gratuitous concurrence, or more plainly put, the inclination of many Aboriginal patients to agree automatically with their clinicians without concern for accuracy, is a common practice which can seriously devalue the treatment and care plan outcome of the PPIE. This is a product of weak cross-cultural communications, sometimes arising due to the 'power distance' between patient and clinician. As we build the ontologies we are breaking down and analysing the most distinct differences between AE and SE discourse. In this process we are identifying the interplay requirements of the semantic and syntactic layers and ontology filters that will effectively intercept enquiry and response content and propose alternate meanings of questions asked and information offered in both directions during or in support of the PPIE.

Semantics and syntax deliberations may occur within or in consequence of many different facets of ontologies supporting AE-centred PPIE. At Figure 2 is the Linguistics parent class, with the '*Lexicosemantics*'

(Lexicon Semantics) subclass and *Synonyms* properties; and at Figure 3 is the '*QuantificationSpecification*' parent class with a number of subclasses and different properties that describe how the Aboriginal world view quantifies and qualifies such things as Temporal which references time and its relationship with events; and Measurement which also expresses distance, dimensions and size in relational terms.
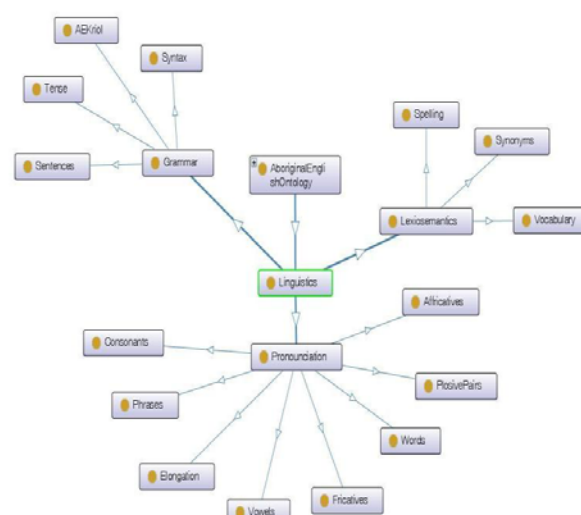
Figure 2: Aboriginal English linguistics
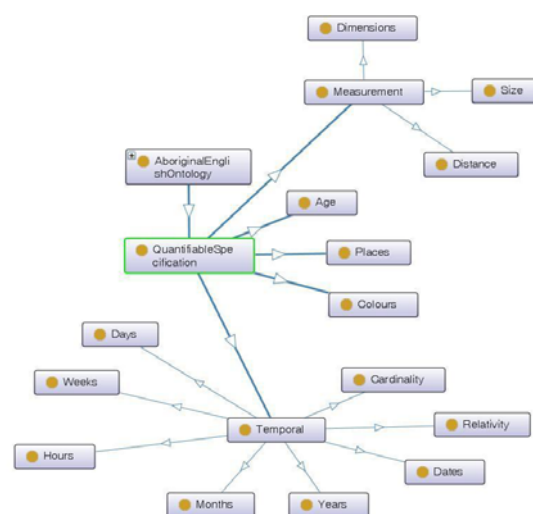
Figure 3: Quantifiable specification

These characteristics do not align comfortably or translate easily with Standard English forms of explanation or questions about quantity. In particular chrono-linear descriptions familiar to westernized clinicians have little meaning in a culture for which there is no beginning and no end.

Table 1 is a sampling illustration of the nature of the proposed PPIE arena, explained in Standard English. The table represents preparatory work for pragmatic contextual translation to AE comprising (a) selected data from Diabetes Management in General Practice Guidelines for Type 2 Diabetes 2009/10 (DIABETES AUSTRALIA, 2009); and (b) research-based commentary and guidance on ambiguity, misinterpretation and issues (Lowell, 1998). The latter is found in AE educational

literature (Malcolm et al., 1999) and will be expanded from the work of the Aboriginal focus groups.

Sample heading T2DM Clinical Guidelines (RACGP /Diabetes Australia)	Preparatory advice/commentary for conversion (code-switching) to AE ontology supported PPIE
Note: *General Form of patient interaction is Standard Australian English conversational and written text, minimal diagrams*	• *Aboriginal English (AE) PPIE will use little to no formal/technical words* • *will significantly reduce written text* • *interface will optimize artwork/diagrams, audio-visual media*
Initial assessment Note: Westemized clinical assessment lacks cultural empathy.	• Use empathetic AE expressions to secure trust. • **Direct and multiple choice questions are inappropriate and commonly ineffectual.** • Adapt to 'Yarning' interactive mode. • Anticipate gratuitous concurrence. • Establish patient-preferred quantifying, • Use AE ontology PPIE system
Plan of continuing care Note: Westemized assumptions unsafe. Patients may overtly or covertly prefer ignorance of their condition.	• AE Two-Way to assess disadvantaged circumstances, home carer environment
Referral :Expect anxiety, reluctance to travel	• Anticipate need to explain in basic words what the referral means.
Counselling the person with diabetes (Includes Initial management; Nutrition; Physical activity)	• Aboriginal Nurse or other Aboriginal Health Worker (AHW) presence/ contribution important. • Use AE PPIE, avoid posing question:"Do you understand?"
Self-monitoring; Medical monitoring; quarterly & annual reviews. Extreme care with terminology.	Explanations for the patient are important. AE PPIE will simplify by avoiding technical terms.
Medication + Problems with medications.	• Identify/ intercept/remedy gratuitous concurrence and non-adherence to recommended medication routines. • AE PPIE uses simplified dosage advice including, side-effects, consequences of non-compliance.
Oral hypoglycaemic agents Insulin treatment; Insulin delivery; Insulins available	• As above. • Correlate with pathology. • AE PPIE to explain 'what' 'how' and 'why'.
Hyperglycaemic emergencies; Diabetic ketoacidosis; Hyperosmolar non-ketotic coma; Factors complicating management	• Use AE PPIE system to explain. • Facilitate sharing clinically preferred response actions with carers/family.
Macrovascular disease; Hypertension; Dyslipidaemia; Renal damage; Eye damage; Foot problems; Neuropathy	• Use AE PPIE system to simplify explanation. • facilitate sharing implications with carers and family. • Anticipate misunderstanding phrases in AE PPIE yarning formats to intercept/remedy.
Diabetes and reproductive health ; Pregnancy Gestational diabetes	• Consult with Aboriginal Nurse or Aboriginal Health Worker for bi-directional accuracy • Anticipate gender sensitivities

Table 1: Sampling of T2DM guideline processes for preparatory mapping to AE ontology

4 Evaluation

The Aboriginal focus group engagements will extend to a series of reviews of the core and stem content of the phraseology emanating from the earlier of the research and focus group findings. This will seek to test and establish reliability of both commonly used every day speech with cultural grammatical constructs; and circumstance-driven embedding of clinical language AE equivalents. Efficacy of the ontological framework will also be tested refined and enabled through scrutiny by and advice from primary care General Practitioners and clinical health expertise available through ongoing academic supervision. Review findings will be examined for pertinence to the applications and interface concepts, and options will be considered where it proves necessary to orientate the intended outcome through ontology changes. In this event, significant (versus marginal) change will be revisited for evaluation by the sources described above. In the context of interface acceptance, e.g. style and ergonomic factors, examples of touch screen and surface computing models will be displayed for viewing and discussion in the latter focus group sessions.

Our work has accordingly taken us toward development of a system whereby ICT will offer intelligent support for communications regardless of location in what we term the Patient-Practitioner Interview Encounter (PPIE). We recognize that perception of 'patient empowerment' may unintentionally suggest a severe shift away from established and reliable T2DM clinical assessment and care management. The emphasis however is concentrated on bi-directional communications assistance that will not only enhance the Aboriginal patient opportunity to contribute to the PPIE, but will reinforce the value of and reciprocal respect for, sound clinical practice. For this we rely upon accredited Australian T2DM guideline sources, notably from the Royal Australian College of General Practitioners and Diabetes Australia. This is further supported by accredited sources of Diabetes Education and Aboriginal Nurse Training, as well as WA state government education specialists in Aboriginal literacy.

5 References

INTERNATIONAL DIABETES INSTITUTE. 2001. The Australian Diabetes, Obesity and Lifestyle Study 'Diabesity & Associated Disorders in Australia – 2000 The Accelerating Epidemic' International Diabetes Institute, Canberra

ENDOCRINE HEALTH NETWORK. 2008. Diabetes Model of Care. *In:* DEPARTMENT OF HEALTH WESTERN AUSTRALIA, ENDOCRINE HEALTH NETWORK (ed.). Perth.

DIABETES WA. 2009. Diabetes in Western Australia. *In:* DIABETES WA AND EPIDEMIOLOGY BRANCH, DEPARTMENT OF HEALTH, WESTERN AUSTRALIA (ed.). Perth.

DIABETES AUSTRALIA. 2009. Diabetes Management in General Practice - Guidelines for Type 2 Diabetes. Diabetes Australia and The Royal Australian College of General Practitioners.

ABS. 2010. *4704.0 - The Health and Welfare of Australia's Aboriginal and Torres Strait Islander Peoples, 2010* [Online]. Australian Bureau of Statistics. Available: http://www.abs.gov.au/AUSSTATS/abs@.nsf/lookup/4704.0Chapter3002010 [Accessed 27 July 2010].

HealthInfoNet. 2010. *Summary of Australian Indigenous health, 2009* [Online]. Mount Lawley WA: Australian Indigenous HealthInfoNet Available: http://www.healthinfonet.ecu.edu.au/health-facts/summary [Accessed 27 July 2010].

WHO. 2009. *Chronic Diseases* [Online]. World Health Organization. Available: http://www.who.int/topics/chronic_diseases/en/ [Accessed 22 July 2009].

AULD, G. 2002. What Can We Say about 112,000 Taps on a Ndjebbana Touch Screen? *The Australian Journal of indigenous Education* 30 1-7.

BARLOW, J. 2009. Creating a Universal Medical Language for WHO Mayo Clinic. *Discovery's Edge.* Online: Mayo Clinic.

BEVERIDGE, M. & FOX, J. 2006. Automatic generation of spoken dialogue from medical plans and

ontologies. *Journal of Biomedical Informatics,* 39**,** 482-489.

CARRILLO, J. E., GREEN, A. R. & BETANCOURT, J. R. 1999. Cross-Cultural Primary Care: A Patient-Based Approach. *Ann Intern Med. ,* 130**,** 829-834.

FISCELLA, K. & EPSTEIN, R. M. 2008. So Much to Do, So Little Time: Care for the Socially Disadvantaged and the 15-Minute Visit. *Arch Intern Med,* 168**,** 1843-1852.

GOH, O. S., DEPICKERE, A., FUNG, C. C. & WONG, K. W. 2007. A Multilevel Natural Language Query Approach for Conversational Agent Systems. *IAENG International Journal of Computer Science,* 33.

KAGAWA-SINGER, M. & KASSIM-LAKHA, S. 2003. A strategy to reduce cross-cultural miscommunication and increase the likelihood of improving health outcomes. *Acad Med. 78(6),* 78**,** 577-87.

LAWRENCE, M., DODD, Z., MOHOR, S., DUNN, S., CRESPIGNY, C. D., POWER, C. & MACKEAN, L. 2009. Improving the Patient Journey: Achieving Positive Outcomes for Remote Aboriginal Cardiac Patients. Cooperative Research Centre for Aboriginal Health, 2009

LOWELL, A. 1998. Cultural Knowledge in Aboriginal Health Care. A review of two subprograms of the Cooperative Research Centre for Aboriginal and Tropical Health's Indigenous Health and Education Research program Available: http://www.crcah.org.au/publications/downloads/Communication_and_Cultural.pdf [Accessed 12 November 2009].

M.CRIBBS & K.GLAISTER 2007. 'It's not easy': Caring for Aboriginal clients with diabetes in remote Australia. *Contemporary Nurse* 25 163-172.

MALCOLM, I. G., HAIG, Y., KÖNIGSBERG, P., ROCHECOUSTE, J., COLLARD, G., HILL, A. & CAHILL, R. 1999. *Two-Way English: Towards More User-Friendly Education for Speakers of Aboriginal English.,* Mount Lawley, Centre for Applied Language and Literacy Research, Edith Cowan University and Education Department of W.A.

P.VAN AST & A.LARSON 2007. Supporting Rural Carers Through Telehealth. *Rural and Remote Health (The International Journal of Rural and Remote Health Research, Education Practice and Policy)* 7:623

SECKIN, G. 2010. Cyber patients surfing the medical web: Computer-mediated medical knowledge and perceived benefits. *Comput. Hum. Behav.,* 26**,** 1694-1700.

SHAHID, S., FINN, L. D. & THOMPSON, S. C. 2009. Barriers to participation of Aboriginal people in cancer care: communication in the hospital setting. *The Medical Journal of Australia,* 190 574-579.

TRUDGEN, R. 2000. *Why Warriors Lie Down and Die,* Parap, NT, Aboriginal Resource and Development Services Inc.

Web-Based Database and SMS to Facilitate Healthcare Medical Emergency

SHIHAB A. HAMEED, SHAHINA SHABNAM BT MOHD SHARIFUDEEN, NUR HAFIZAH CHEK NUH , NUR HUDA BT SALIM, AISHA HASSAN, OTHMAN KHALIFA

Electrical and Computer Engineering Department
Faculty of Engineering, IIUM University, 50728 KL
Kuala Lumpur, Malaysia
shihab@iiu.edu.my

Abstract

Healthcare and medical emergency are essential systems in human life; so that many countries work toward having it. Investigation of Malaysian case shows that it suffer from locality, missing of unified electronic medical record EMR and lack of utilizing Internet, multimedia, wireless and real time technologies. Other drawbacks include: Difficulty in searching and viewing up to date records for patient, doctor, hospital, and drug, since many of such records are still kept in filing cabinet. This leads to difficulty in communication, hard to manage and exchange patient data between various medical units. The fully computerizing and combining of such medical systems will lead to produce a Novel Integrated Healthcare Medical Emergency Model (IHMEM). IHMEM includes three main parts: The web based database, the intelligent agent, and the mobility. This paper focuses on developing an interactive web-based database with unified EMR as well as using SMS facilities, where all hospitals, healthcare and emergency centers can view the patient record simultaneously, exchanging, managing and collaborate on sharing resources between medical units. A prototype for this work is build and sample of implementation results for interactive database and SMS facilities are shown. IHMEM can be easily customized to be suitable for other countries.

Keywords: web-based database; healthcare; medical emergency; intelligent agent; SMS

1. Introduction

Internet plays a significant role in connecting all the participants in the health community. It is used to control remote medical equipments, communication between parties such as patients and doctors, search for needed information, transferring text, graphics, audio and video files as well as supporting collaboration in real-time [1].

In the web-based multimedia environment, the images for medical model can be categorized into different types: magnetic resonance (MR), computerized topography (CT), X-ray, electrocardiograms (ECG) among others as well as medical information in forms of charts, graphs and others. These images could be loaded electronically with digital devices into the patient medical information. Thus, this would prevent the patient's medical images from damage or lost. Besides, it would be much comfortable for both patient and doctor, where, a patient may go for treatment in other medical center without carrying the medical report. The doctors could also view the patient health related images for further clarification without re-examine the patient [2].

A mobile phone and other mobility devices can be used to link to the model. The wireless application protocol (WAP) is used to standardize the way mobility devices associated for Internet access, including e-mail and the World Wide Web [3].

Survey study for the current Malaysian healthcare, medical and emergency systems are still suffer from locality, paper-based and lack of utilizing Internet, multimedia, wireless and real time technologies. More details on drawbacks are shown in many related publications [4][5][6][7][8].

On other hand; an emergency system reduces the risk of an emergency case to the health and safety of persons and valuables by providing an effective means of communication with relevant authorities, safety guidelines and measures to be taken in an emergency situation [9]. Several of these systems are normally rely on the existing telephone and other communications infrastructure via operators and service personnel; therefore they suffer from several drawbacks [10].

Many hospitals and emergency centers are not efficient enough because the big number of emergency cases, which is not easy to be handled. In an emergency department, most likely a nurse will determine the severity of the wound and check patient's vital signs such as temperature, blood pressure and heart rate. Additional personal information and medical history have to be obtained. In case patient's information is stored in another clinic or hospital they have to be obtained. Unless the patient has brought the personal file along, getting the required information will slow down the process. An emergency physician will have to examine the patient. In some cases, the patient needs great attention and sometimes the surgery room or other devices will be needed. All these will have to be arranged in timely manner, which in general is not the case.

2. Existing Medical Related Applications

Telemedicine can be defined as the use of audio, video, and other telecommunications and electronic information processing technologies to provide health services or assist health care personnel at distant sites [11]. Nowadays the evolution of wireless communication means enables telemedicine systems to operate across the world, increasing telemedicine benefits, applications, and services. The following are sample of projects that have been developed in the field of telemedicine and communication.

2.1 **Momeda (Mobile Medical Data)**: is a demonstrator that can be used from a PDA (Personal Digital Assistant) to access electronic patient record data and provide it to the consulting physician. Diagnostic information such as radiological images as well as text and laboratory data is transmitted to a wireless pocket-size terminal in a user-friendly multimedia format using Web-approach.[12] It allows patients to access customized disease-specific information material that enables them to fully understand in a simple and constructive form what their medical problem is, what the planned procedures are, what lifestyle they should follow during and after their hospitalization, thus becoming more qualified partners in the recovery process.[13]

2.2 **The Ambulance project**: was developed by the national university of Athens. They develop a portable emergency telemedicine device that supports real time transmission of critical bio-signals as well as still images of the patients using GSM link. [14]

2.3 **Emergency-112**: is an extension of the ambulance project. It targeted to: reduces treatment times, improve medical diagnosis, and reduce costs by developing an integrated portable medical device for Emergency Telemedicine. The transmission of critical bio-signals (ECG, BP, HR, SpO2, and Temperature) and images to an Emergency call centre enables physicians to direct pre-hospital care more effectively, improving patient outcomes and reducing mortality. Networking links to medical information databases, Hospital Information Systems, and Inter-hospital links are also provided to maximize information available to consulting physicians. The Emergency-112 system has been used successfully since 1998 in three European Countries (Greece, Italy, and Cyprus). Nevertheless, as the above projects mainly use a slow GSM link (9.6kbps), it cannot incorporate video along its transmission nor can it support high resolution imaging [15].

2.4 **Multimedia Telemedicine System (MTS)**: is a client/server architecture that uses TCP/IP over the Internet. Doctor with patient and doctor can communicate each other by exchanging real-time data including audio, video and instant message (IM), and non-real-time data, including vital sign signals, radiological images with DICOM 3.0, file, bio-signal, bio-data etc. [16]

2.5 **Project E-vita**: is a commercial browser based electronic health record system with a low bandwidth requirement. Patient clinical encounter history with nurses, doctors and other healthcare professionals from multiple agencies can all be recorded in one easy to use browser based patient record. [17]

3. Integrated Healthcare Medical Emergency Model

L.R study and Investigation of the current situation of healthcare and medical related systems in Malaysia have several weaknesses. To overcome it we have to develop an Integrated Healthcare Medical Emergency Model. Based on requirements' elicitation, analysis and specifications; IHMEM developed with interactive web-based multimedia environment and mobility technology. The model provides an integrated healthcare and medical database, which can provide stakeholders with related medical information. The registered users can log into the system to access or provide medical information based on their accessing privilege. The medical information can be stored in a variety of multimedia forms such as video, audio, pictures and text. For example, in addition to text description of patients' historical medical information, graphic images such as X-rays or video files of doctors' discussion about the disease can also be saved in patients' record.

The model have the capabilities for finding the patient location based and suggest the nearest emergency center, arrange all necessary related patient information to be ready for the physician when the patient arrives, assigning a doctor to the patient based on the availability of the doctors and list all necessary requirements (if any) such as special devices or surgery room. "Fig.1" shows the IHMEM architecture.
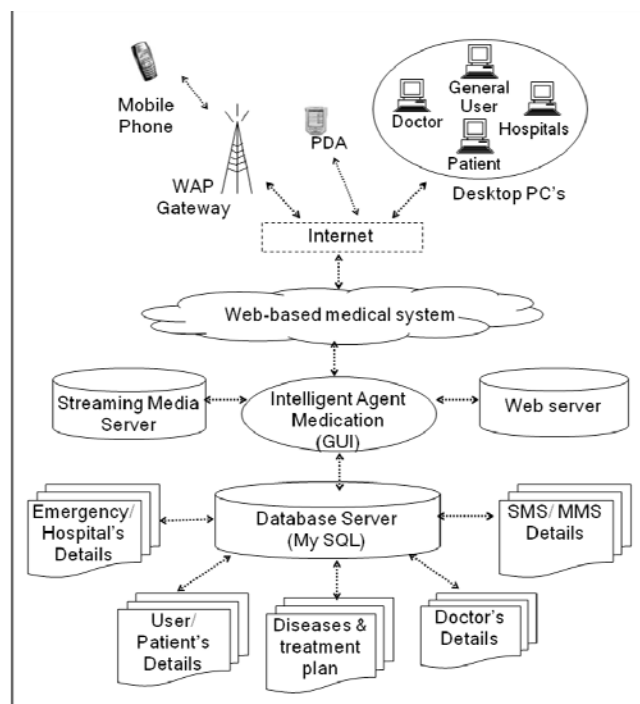

Figure 1: Overall IHMEM architecture

The model is an open cross-platform web-based real-time client-server environment with multiple language capabilities. The client system is loaded with multi-form interfaces, database access functionalities, and multimedia information processing and manipulation tools. The system provides mechanisms for exchange of image files, shared discussion lists, textual information exchange, access to images and data exported from local data bases, voice and video transmission.

The scripting language used to build the model is PHP and MYSQL database. MySQL is a true multi-user, multi-threaded SQL database server. PHP is an HTML-embedded scripting language. The goal of PHP is to allow web developers to write dynamically generated pages. By implementing MySQL and PHP together, one can design a functional web-based database quickly [18]. The main components of the IHMEM include:

3.1 Database: this is a fundamental part of the system. It stores all important and detailed information about general users, emergency authorities, doctors, patients, hospitals and emergency centers, places or locations and events within the area of implementation, set of prerecorded SMS and MMS, which are suitable for different emergency and guidance cases. In addition, the database supports real-time multimedia [19] [20].

3.2 Web Interface: the interface is simple, user friendly and requires little input from the user, mostly in the form of choices. It is based on Web 2.0 technology and has multiple language features.

3.3 Web Server: will listens for requests from Web browsers and upon receiving a request for a file sends it back to the browser. It will host the program and control information for the system

3.4 Telephony Server: will act as a Private Branch Exchange using the open source Asterisk PBX [21]. Users are able to call the PBX number and will be prompted with an interactive voice response (IVR).

3.5 The SMS server: will continually listen for incoming SMS, process and pass them to the Agent. The agent will respond accordingly and the SMS server will deliver the response back to the user.

3.6 Streaming Media Server: this is a dedicated Streaming Server for streaming multimedia to the stakeholders. It provides high quality media, effective bandwidth utilization, and supports detailed reporting and multi-stream multimedia for larger numbers of users [22].

For the database part, the IHMEM utilized the 3-tier architecture including browser, web server and MySQL database server as shown in "Fig.2". The client sides of the model are the system administrator application, medical center (MC) administrator application, doctor and nurse application, or any server client.
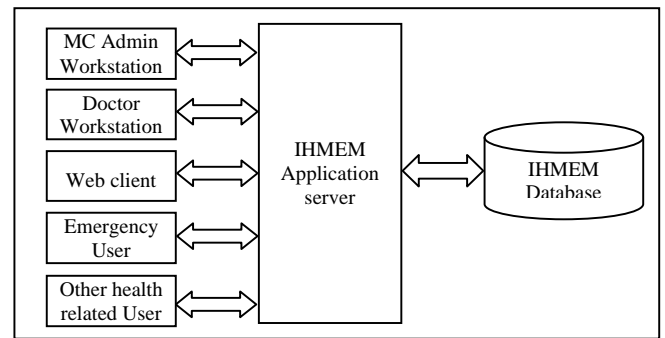

Figure 2: IHMEM system database architecture

The database stores all the medical center information, doctors' record and patient records. The database records are mapped in the IHMEM system application server.

4. IHMEM Implementation (Interface and Database Design Module)

A prototype for IHMEM is developed and sample of the implementation of the main interface and database architecture are shown in this section. A prototype of the suggested model with some of the above mentioned features have been implemented. They include client application, administrator application, doctor application and mobility access system application.

Client application: is a way for the user to go through the system. "Fig.3" shows the architecture of the client application. Database modification is done through the Data Manager layer. User login is done on the same page, but from different table in the database. The data is retrieve from the database to be send to the interface (back to user). All the data of the users are stored in the database. The communication is done at the Data Manager layer.
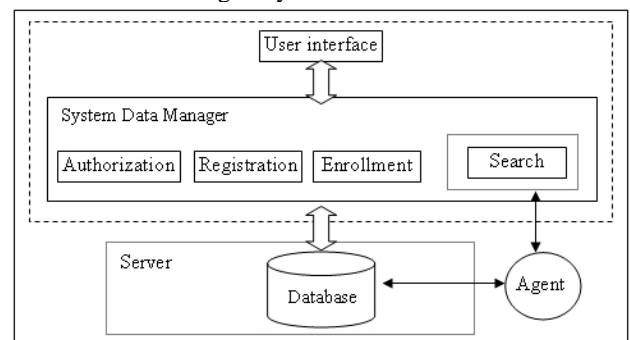

Figure 3: Client application architecture

Administrator application; Administrator has the authority to modify all the data and the system. Admin page is important to manage data and authentication of the patients and also doctor. The responsibilities of the administrator page functions are facilitate: addition and updating of patients and medical centre with its staff, add/update the working schedules of the staff /doctor.

Staff / doctor application: The responsibilities of the Staff page functions are: Accessed past medical report of the patient; the summary of the patient report which is very important for better treatment is available. The staff could record the disease and the appropriate medicine after the consulting or treat the patient; this will be recorded in patient's database for future references. Staff could record any new disease information and the suitable treatment plan. All the records are update after patient's visit

Patient Application: It is responsible for: Viewing personal and contact information, having the information of the patient next treatment or checkup, searching function for the medicine information, doctor and medical centre.

Database is created using MySQL codes. The database will store all the information of patients, staff and medical centre. Patient is relate with staff in terms of staff manages data for patients. For staff and medical centre, the staff relate with medical centre as a worker in that medical center. "Fig. 4" shows the entity relationship for the medical emergency model.

Medical Emergency model can be accessed either through the server in network system or local-host in personal computer. The model is divided into two main parts. The first part is called administration page; only authorize user have permission to access this page by entering username and password. The second part called main page so all users can access it. When user accesses the second part; there are several options for user to choose which one they want the model to do it first. The main page will displayed home, about us, health care centre and search as shown in "Fig.5".

Searching function can be used by user to find list of: doctors, medicines, diseases, medical centre, and health tips by inserting the criteria needed. The flow of the model is shown in "Fig.5".
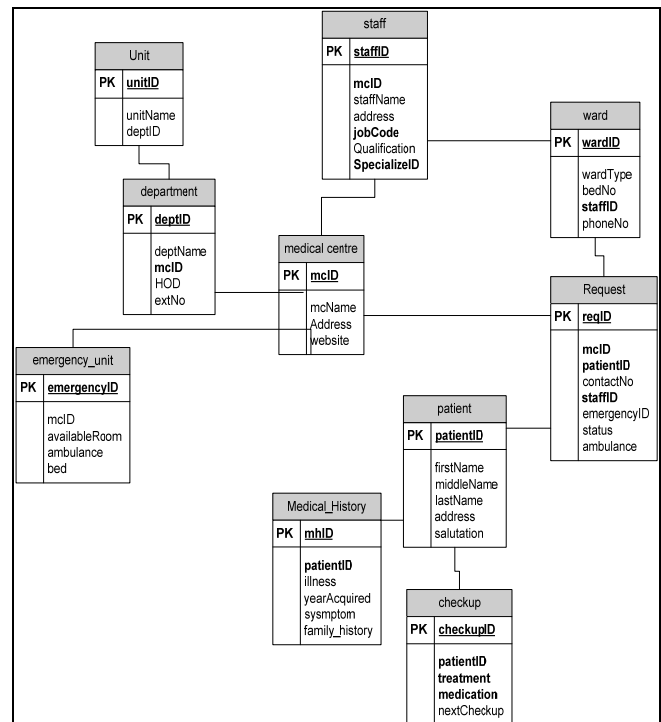

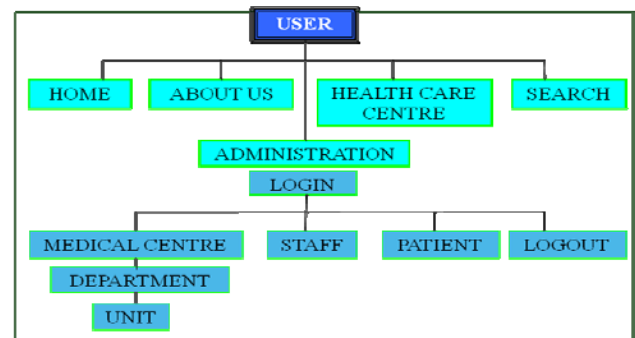Figure 4 Entity Relationship Diagram for IHMEM Model

Figure 5 IHMEN web flow

5. Database Implementation Results

In this section we will present sample results for implementing IHMEM model. IHMEM model provides the registration interface for different class of users to register then view different data in database. Administrator will register the staff and their username and password will be given by the admin. The following example is for registration of new patient as shown in "Fig.6". But, admin can only register new staff and new medical centre and only staff can register new patient.

Registered user can use the graphical user interface to log into the IHMEM by entering username and password; user needs to choose in which group they belong to, staff, patient or admin. "Fig.7" shows the login interface. Each user will be differentiating by the group and during the login after validation step; each user will be directed to the corresponding page based on their group.
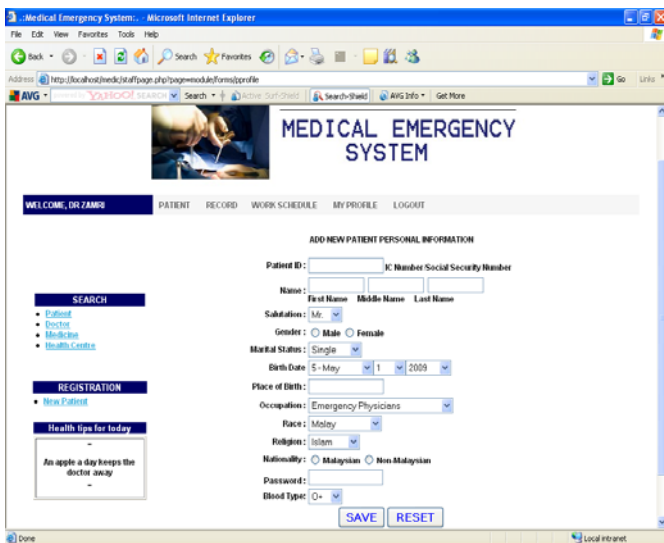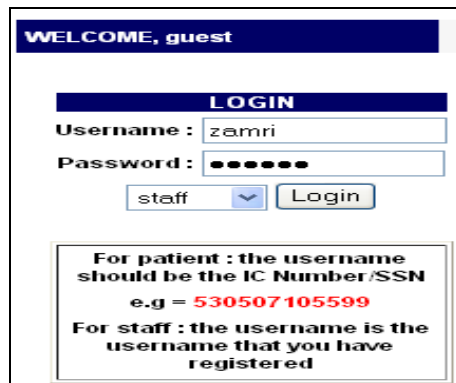
Staff has the authority to view patient's detail. In staff page, we can view the patient's detail as shown in "Fig.9" for a list of patients. If we click on the patient name; the full detail of that patient will be displayed.
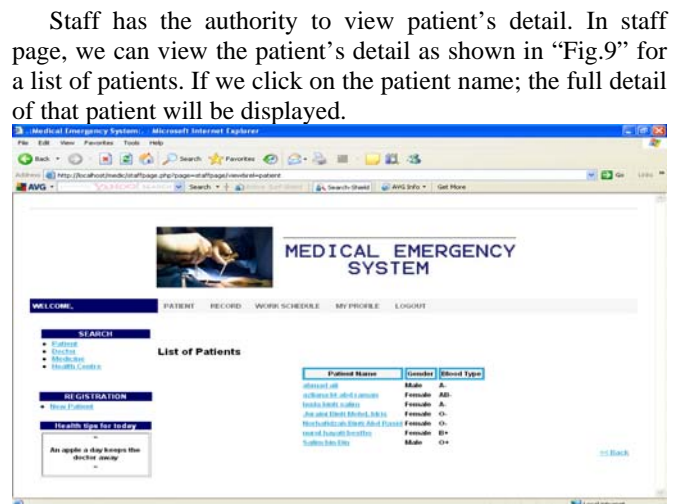

Figure 6 New Patient Registrations

Figure 9: List of Patients

Or, for specific patient, staff can do searching by inserting patient's IC number and click Submit. Then the server will sent the input to the database and search for patient with the given input. Sample searching result is shown in "Fig.10".
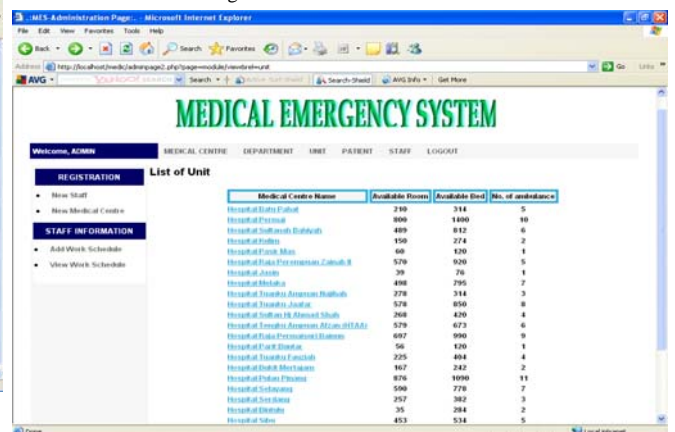

Figure 7 Login Page

Figure 10: Patient search result
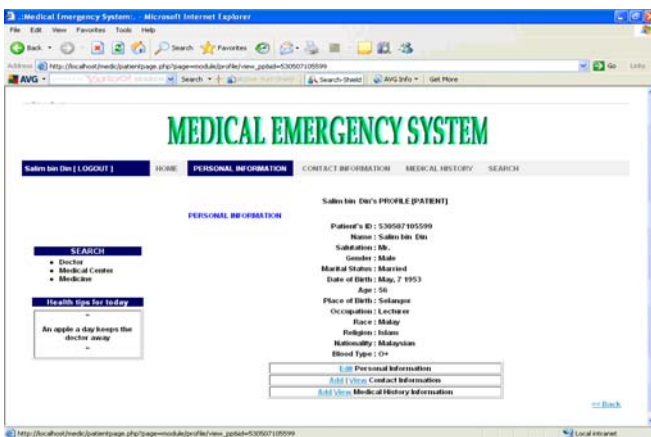
Figure 8: Patient personal information

Figure 11 list of Medical Centre with emergency unit

After login validation user in group patient will be directed to the patient page. In this page, patient can view their own personal and contact detail information. "Fig.8" shows details for patient personal information.

The model offers searching the availability of the emergency unit. With this information, user can know whether the medical centre has emergency unit or not, the number of

available room, bed and ambulance. "Fig.11" shows available medical center and emergency units.

The doctors work schedules is very important to identify the availability of doctor. The system has the capability to add/update and delete the work schedule. The administrator is responsible for add/update their own respective doctors' work schedules. "Fig.12" shows the list of staff with their work schedule

Figure 12 list of staff with their work schedule

6. SMS Module

One of the main facilities of the mobility is the using of SMS and MMS. In this paper we will focus on using the SMS facility within IHMEM. SMS module provides three main features: locate nearest healthcare center, healthcare center search and SMS broadcasting. Non-functional requirements related to this module include the following:

- The features of this module must be available only to devices that can send or receive SMS messages

- No authentication is required for:
  - locating nearest health care center requests
  - healthcare center search requests
  - SMS user registration for SMS broadcast feature

SMS module consists of two actors and three use cases. User initiates 'locate nearest health care center' and 'health care center search' use cases, while staff initiates 'SMS broadcast' use case.

6.1 Locate Nearest Hhealthcare Center by SMS

Locating the nearest health care center by using SMS is an essential feature of the system. This feature comes in handy when a patient is in need of medical attention but is not aware of the nearest health care center. Responsibilities of this module unit include:

- Handling receiving of SMS requests for nearest healthcare center

- Locating nearest healthcare center based on the provided input

- Sending SMS for nearest healthcare center information to requester

As seen in "Fig. 13" the module needs to communicate with Google Geocoding service and local hospital database. In our model all the SMS requests are received by the GSM modem. In the telephony server, a daemon continuously running in the background listens for incoming SMS requests. Once a request is received by the GSM modem, it is placed by the daemon to a queue for processing.

Once the nearest healthcare center is found, its name, address and contact information are retrieved and send by SMS to the requester.
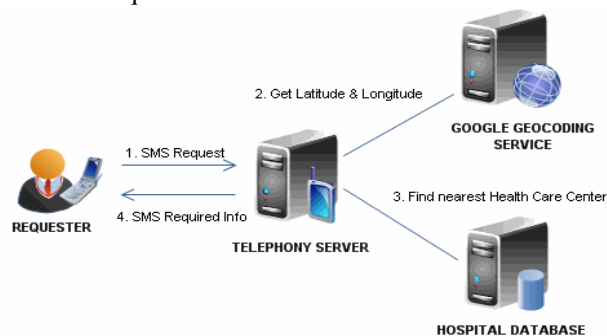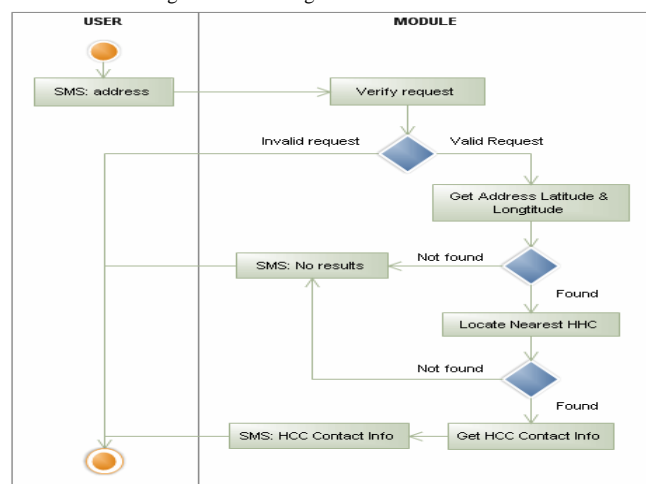

Figure 13: Locating nearest healthcare center

Figure 14: Activity diagram for Locate nearest healthcare center

"Fig.14" displays the UML activity diagram for locating the nearest health care center. It is divided into two swim lanes indicating the activities completed by the requester and the module. The activity diagram furthermore includes checking for invalid input and handling the situations when the geographical location data cannot be retrieved or the nearest health care center is not found.

For emergencies starting out of medical care, it is essential to know the location of the nearest healthcare center to get assistance from qualified personnel using SMS. User is required to send an SMS to the system providing partial or full address of the current location. The system will respond back by sending an SMS message providing the nearest health care center name, full address and contact numbers.

Since this model using a single GSM modem to handle the SMS and MMS requests; so that some keywords are used to

differentiate between different user requests. To search for the nearest health care center the keyword **SNH** is used. The message format to request for the nearest health care center is:

SNH *<current_location_address>*

The SMS module will reply back to the requester with an SMS message containing the nearest health care center relevant information. "Fig.15a" shows SMS message request for the nearest health care center. The SMS module will locate the nearest health care center and return the SMS message shown in "Fig.15b", which contains the relevant search results.
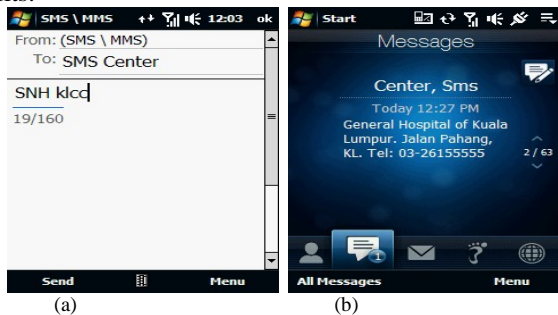

(a)   (b)

Figure 15: Finding nearest healthcare center by SMS (a) Request SMS message (b) Received SMS message

6.2 Healthcare Center Search by SMS

SMS search module allows users to search for health care centers using short message service (SMS) platform. Users send queries to a provided universal short code (which in our model is a mobile phone number) and receive results on their phones. As shown in "Fig.16", this module requires a single input and provides the output based on the search string.

The following is the list of input and output:

- Input: Health care center partial or full name (mandatory)

- Output: SMS requester top two search results.

- Results include healthcare center name, address and contact information.

Unlike locating nearest health care center module, the search module doesn't require a communication session with the Google Geocode service. It only queries the local health care center database.
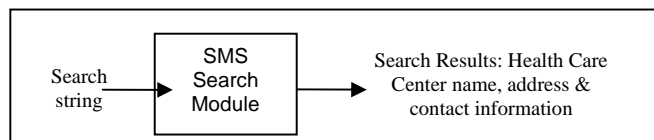

Figure 16: Input and output for SMS search module

SMS search module consists of a daemon running constantly in the background listening for incoming SMS search requests. Once the search requests are received by the GSM modem, they are placed into a queue by the daemon for processing. "Fig.17" shows the flow and components of the healthcare / emergency center SMS search.

Figure 17: Health care center SMS search

In fact the search may return more than one result. Thus, in our model only top two results are sent to the user. However, this can be easily changed by editing the module configuration file. Since the length of the text of an SMS is limited to 160 characters, each result is send as a separate message. Each SMS message contains health care center name, address and contact information. "Fig.18" displays the UML activity diagram for a single SMS request.
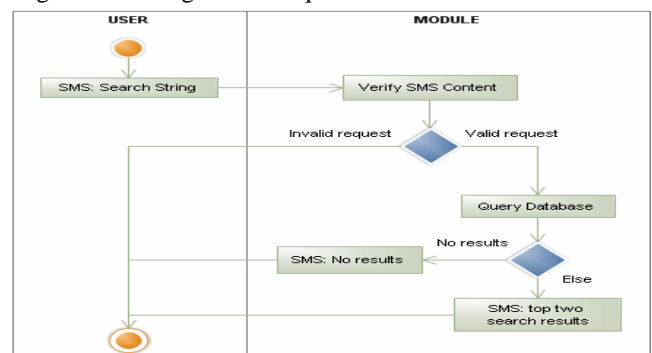

Figure 18: Activity diagram for the Health care center search

The SMS module allows the user to search for hospitals by using their mobile phones. Users send SMS search queries to the provided short code (GSM Modem number) and the SMS module will return the search results. When searching for a hospital, the search might return more than one result. Only the top two search results will be send to the user each in different SMS messages. The request SMS format is:

SH *<hospital_name>*

(a)   (b)

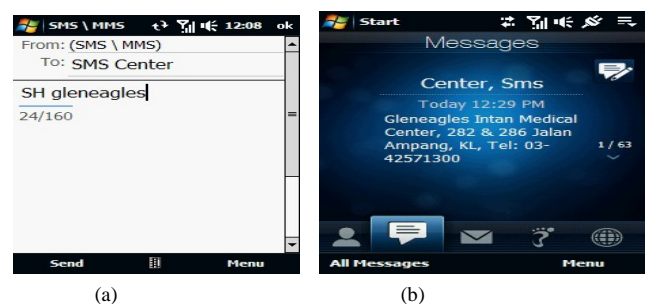Figure 19: Hospital search by SMS (a) Request SMS message (b) Received message

"Fig.19a" shows a hospital search request SMS message. The user is searching for a hospital that contains the keyword *Gleneagles* in the name. After the SMS is received by the SMS module, it will find and SMS back the top two search results. In this case there is only one search result returned that is sent back to the user as shown in "Fig.19b".

Locating the nearest health care center using SMS is a very useful feature. It offers to the user great flexibility. However, SMS cannot be considered as a real time communication medium. The time that the SMS message reaches the server or the user will depend on the condition of the telephony network. However, telecommunication companies offer a faster SMS service for a certain fee. This will speed up the SMS communication between the user and the system.

6.3 SMS broadcast

SMS is a very efficient way to communicate short important information to the public. It can be efficiently used to send advices, preventions and treatment messages to patients related to different diseases. The following is a list of input and expected output of the SMS notification module:

Input:

- message content (mandatory)
- destination mobile numbers (mandatory)
- date and time when the message is to be send (mandatory)

Expected output:

- SMS broadcast the specified content to the provided destination numbers at the specified time.

Destination mobile number is one of the input requirements of the module. There are two ways to get the destination mobile numbers. The flow of activities involved in the sending of SMS notification is revealed in "Fig.20". The module carries out the following activities:

- get SMS content
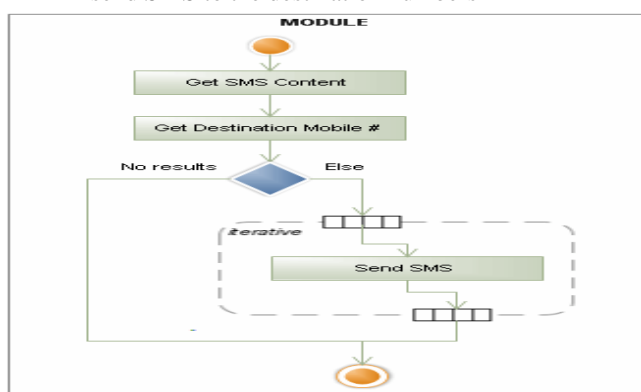- get destination mobile numbers
- send SMS to the destination numbers

Figure 20: Activity diagram for the SMS notification

As seen in "Fig.20" all activities are carried out by the module. The module consists of a daemon continuously running on the background checking if there is any scheduled message to be sent by SMS. If a message is found the module initiates a connection to the database and gets the content of the message. Subsequently, the list of the destination mobile numbers is pulled out from the database and placed in an array. After that the module will connect to the GSM modem and send the message to all the destination mobile numbers.
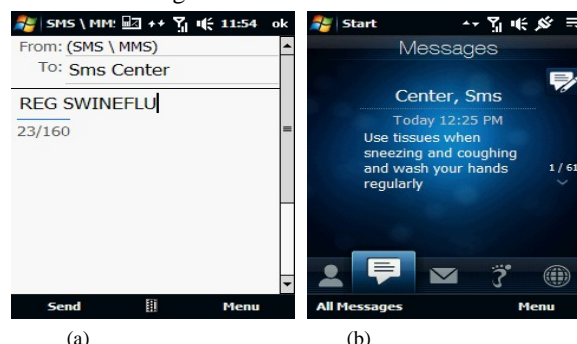

Figure 21: (a) Registration SMS message (b) Sample SMS message received by the system

SMS is a very efficient way to communicate short important information to the public. SMS can be efficiently used to send advices, preventions and treatment messages to patients related to different diseases. SMS can be used to send prevention messages, which in return can help the spreading of the disease. "Fig.21a" and "Fig. 21b" show how it works:

- User registers to receive SMS messages by sending an SMS message to the provided short code (in our model the phone number of our GMS modem) with the message: REG SWINEFLU.
- The system will send to the user SMS messages containing prevention or other useful information. These SMS messages can be sent by the administrator by simply filling a form and the system will automatically send the SMS message to all registered users.

7. Conclusion

Locality, lack of: mobility and multimedia as well as missing of unified database and electronic medical record are main drawbacks for current medical related systems in Malaysia. We develop an Integrated Healthcare Medical Emergency Model (IHMEM) to overcome such problems.

This paper focuses on developing and implementing a prototype for an interactive web-based database with unified EMR as well as using SMS facilities, where all hospitals, healthcare and emergency centers can view the patient record simultaneously, exchanging, managing and collaborate on sharing resources between medical units. Sample of the results are shown in this paper which reflect some of the capability of this model in offering insertion, deletion, updating and searching for different related data.

The results show the important of this model in serving different stakeholders such as doctor, nurse, patients in their daily activity. IHMEM is a very helpful user friendly tool for patients and other users especially in emergency cases since it offers information about medical centers and available emergency units so the patient can go directly to the suitable and nearest one.

Results of implementing SMS module shows sample SMS facilities in searching and locating for a suitable hospital as well as finding the nearest hospital.

Although this study and model build based on the Malaysia case but it is also suitable to be used for other countries with little adjustment on data and functional description.

8. References

[1] National research Council, *Networking Health: Prescription for the Internet*, National Academy Press, Washington, 2000

[2] Shihab A. Hameed, Shahina Shabnam, "Integrated Medical and Emergency System", ICCCE06, Malaysia 2006

[3] Shihab A. Hameed, Shahina Shabnam, "An Intelligent Agent-Based Medication and Emergency System", ICTTA06, Syria 2006

[4] Leslie Versweyveld, *Virtual Medical Worlds*, http://www.hoise.com/vmw/02/articles/vmw/LV-VM-03-02-33.html

[5] Warner MD, Malaysian Medical Matrix: Telemedicine in the age of Multimedia Super Corridor, http://www.pulsar.org/febweb/papers/m3web.htm

[6] Materia Medica Malaysiana, Putrajaya hospital showcases latest eHealth service,http://medicine.com.my/mblog/2004/08/putrajaya-hospital-showcases-latest.html

[7] Mark H. Spohr, Healthcare Informatics Online, Malaysia's 21st Century Goal: IT-Based Health Gains For One And All, http://www.healthcare-informatics.com/issues/2000/04_00/international.htm

[8] Loh Foon Fong, *Stuck In The Cyber Web*, www2.jaring.my/pcdom/Stuck%20in%20the%20cyber%20web.pdf

[9] A Guide to Setting up a Workplace Safety and Health Program, http://www.safemanitoba.com/pdf/wshprogram_setup_guide.pdf

[10] Shihab A. Hameed, B.A. Aliyu, "WIMAX Wireless Technology to enhance an Integrated Emergency and Guidance System", Accepted in ISM06, UAE 2006

[11] http://www.acponline.org/computer/telemedicine/glossary.htm

[12] *The MOMEDA project.* http://www.biomed.ntua.gr/ momeda

[13] Pavlopoulos, S. Prentza, A. Kyriacou, E. Marinos, S. Stassis, A. Kalivas, D. Koutsouris, D. Filippatos, G., "A personalized medical information system for patienteducation-MOMEDA", MES/EMBS Conference, 1999. Proceedings of the First Joint Volume 2, Issue , Oct 1999 Page(s):1238

[14] S. Pavlopoulos, E. Kyriacou, A. Berler, S. Dembeyiotis, D. Koutsouris "A novel emergency telemedicine system based on wireless communication technology – AMBULANCE," *IEEE Trans. Inform. Tech. Biomed*. – Special Issue on Emerging Health Telematics Applications in Europe, vol.2, no.4, pp.261-267,1998.

[15] The EMERGENCY 112 project, http://www.biomed.ntua.gr/emergency112, 1998.

[16] Li Ling; Yao Dezhong; Li Jianqing; Li Bin; Wang Ling, "A multimedia telemedicine system", Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the Volume , Issue , 2005 Page(s):3746 - 3748

[17] Project E-vita, http://www.projectevita.com/prod ucts/what_is_pe.aspx

[18] Janine Roe , *Creating a web-based database using MySQL and PHP*, http://taz.cs.wcupa.edu/~jroe/mysqlphp.html

[19] Microsoft Corporation, "Streaming Methods": Web Servers vs Streaming Server", http://www.microsoft.com/windows/windowsmedia/compare/webservvs treamserv.aspx.

[20] Wee, N. C. Real-Time Web Vehicle Classifier with Multimedia Database, A Bachelor degree Thesis, University of Quensland, Oct 2002.

[21] Asterisk PBX, http://www.asterisk.org/

[22] [Ling, S. Real Time Web Vehicle Classifier: Network Configuration, A Bachelor degree Thesis, Dept of Information Technology and Electrical Engineering, University of Quensland, Oct 2002.

Opportunities for Knowledge Management Tools in Clinical Genetic Services

YULONG GU[1], JIM WARREN[1,2], KAREN DAY[2]

[1]Department of Computer Science, [2]School of Population Health, University of Auckland, New Zealand

`ygu029@cs.auckland.ac.nz`

Abstract

This study aims to characterize the challenges in utilizing genetic information in health care and to identify opportunities for genetics knowledge management. Taking a grounded theory approach, we conducted semi-structured interviews with 48 New Zealand genetic services stakeholders. Several themes emerged from the data, including clinicians' lack of knowledge on genetic testing and a lack of support for test result dissemination. Knowledge management tools may have the potential to support key tasks in genetic services delivery, improve knowledge processes, and enhance knowledge networks. Stakeholder comments indicate opportunities for solutions such as decision support systems, electronic referral systems, electronic health record or personal health record systems, data submission and other knowledge processing tools, ontologies, and knowledge networking tools. Implementation of these technologies in clinical genetic services may help unleash the power of genetic information to improve healthcare outcomes and knowledge management performance.

Keywords: knowledge management; human genetics; clinical decision support systems.

1 Introduction

Since the completion of human genome sequencing in 2003, our understanding of the genetic component of human diseases has vastly increased (Engstrom et al., 2005). Human genetic variation knowledge identifies the associations between genetic variations and human conditions. Applying this knowledge, clinical genetic services provide genetic testing and results interpretation (IT Perspectives Inc., 2005). These services may improve patient care because genetic testing can produce valuable information for diagnosis, prognosis, treatment, and avoidance of adverse drug reaction. It is surprising that healthcare has not been transformed by the revolution of (molecular) genetic testing, as predicted by many, including (Bell, 1998, Bhandari et al., 2000, Glasner, 2002, McGuire et al., 2008, Zimmern, 1999, Personalized Medicine Coalition, 2006).

The management of medical genetic test results is far from systematic or efficient in terms of data storage, access, sharing, and usage. These results are not treated as other medical data in health systems. This is because

genetic information has familial and generational implications; it is predictive and the emphasis is not on treatment; its implications are sometimes severe, private and complex, introducing special ethical dilemmas (Kääriäinen et al., 2006). Given these challenges, this study aims to characterize the use of genetic information in health care and to inform how we might improve human genetic variation knowledge management.

Knowledge is a "fluid mix of framed experience, values, contextual information and expert insight that provide(s) a framework for evaluation and incorporating new experiences and information" (Davenport and Prusak, 1997). Accordingly, managing genetics knowledge requires a structured representation of genetic data and a framework in which to evaluate, incorporate, and share new experiences and information from routine genetic testing. It is valuable to support the processing of both explicit and tacit human genetic variation knowledge in order to advance genetics science as well as to apply it in health care. Knowledge management (KM) literature has identified three essential approaches: knowledge stock, knowledge process, and intellectual capital (Nonaka, 1994, Carlsson et al., 1996, Alavi and Leidner, 2001, Schultze and Leidner, 2002, Sambamurthy et al., 2003, Argote et al., 2003, Stankosky, 2005). Based on these theories, this research project aims to explore the principles and goals for human genetic variation knowledge management.

To realize the full potential of personalized medicine, policies must be implemented to appropriately protect the confidentiality, privacy, and security of genetic testing data with regard to access and use (McGuire et al., 2008). In the US, data models are being created to promote incorporation of genetic information in health record storage (Feero et al., 2008) and transfer (Shabo, 2007). New Zealand studies are needed to answer the critical questions around clinical genetic information management principles and frameworks. In this study, we analyse the range of practices in use of information relating to genetic services, and uncover the perceived gaps and opportunities within the existing information technology (IT) environment.

2 Methods

This study takes a grounded theory approach to provide groundwork for refining KM processes in genetic services. Grounded theory is an inductive methodology to generate theories through the emergence of conceptual categories (Glaser and Strauss, 1967, Rhine, 2008). The research data are collected by (1) reviewing relevant literature, policies, and organizational documents; (2) observing and communicating with participants about their roles, activities, and technology satisfaction; (3)

recording stakeholder comments on current or potential issues and challenges, mostly via face-to-face or telephone interviews; and (4) writing notes during observation, data collection, and data analyses. This multi-dimensional approach validates the accuracy of our findings as it triangulates different data sources by examining evidence from the sources and using it to build a coherent justification for themes (Creswell, 2003).

We collected the main data set of the study through semi-structured interviews with 48 participants. The interview questions encourage participants to talk about issues they feel are critical to genetic information management and KM. We used convenience sampling, snowballing, and theoretical sampling to identify participants that represent key stakeholder groups of the clinical genetic services in New Zealand. Their roles include those of patients and family members (numbering 8), healthcare providers (8), genetic counselling professionals (11), genetic testing laboratory scientists (3), governance (3), health IT professionals (4), researchers (5), directors of health institutions providing genetic services (7), and clinical advisors at indirect health services such as health insurance providers (5). Some participants have more than one role in the system, for instance as both clinician and director.

We conducted the data analysis as interviews proceed until data saturation. We use the classic grounded theory approach for inductive framework production from theoretical sampling, open coding, interpretive analysis, and constant comparison (Glaser and Strauss, 1967). We also apply a systematic multiple coding process that developed upon classic grounded theory, including three steps: (1) generating categories of information (open coding), (2) selecting one of the categories and positioning it within a theoretical model (axial coding), and then (3) explicating a story from the interconnection of these categories (selective coding) (Strauss and Corbin, 1998), see also (Creswell, 2003). Through multiple coding over interview data and constant comparison among all data sources, several data categories and themes, as well as one final theory, emerge based on our understanding of the New Zealand genetic service system processes and of stakeholder perspectives. One of these results indicates promising technologies for improving the performance of human genetic variation KM in clinical settings.

3 KM Challenges

A number of topics surface from the interviews concerning genetic information management and knowledge management. In another paper, we reported the identified challenges and concluded that in order to realise the full potential of human genetic variation knowledge for better healthcare outcomes we need to overcome: the challenges of multiple service models, the inhibitors for genetic data sharing, and the barriers to wider use of genetic testing (Gu et al., 2011). In the current paper, we discuss the two themes that present major barriers for managing human genetic variation knowledge in the clinical setting, i.e., the clinician knowledge of genetic testing and the dissemination of genetic test results.

3.1 Clinicians' Lack of Knowledge Regarding Genetic Testing

In an earlier paper (Gu et al., 2010), we reported the gaps in clinician knowledge are a barrier for involving them in genetic services delivery. There is little information available to clinicians regarding genetic test's clinical utility and the service's referring processes. We collect clinicians' need for support in test referral decision-making, referral processing, and patient management. Critical information during these tasks includes disease risk probability, the test's clinical utility, prevalence, penetrance, and cost/benefit, as well as interpretation and limitation of test result. KM literature emphasizes support for knowledge processing, including knowledge storage, application, and reuse (Alavi and Leidner, 2001, Schultze and Leidner, 2002, Sambamurthy et al., 2003, Argote et al., 2003). Accordingly, genetics knowledge processing requires efforts to establish human genetic variation knowledge stocks, such as genetic service protocols and evidence-based clinical guidelines, as well as to facilitate knowledge flows that deliver critical information to clinicians.

3.2 The Lack of Support for Circulating Genetic Information

Our study also highlights a lack of support for distributing genetic test results. A regional genetic service director states that test results are disseminated only when a patient wishes them to be and the patient will benefit from the data sharing. However, there is no health information dissemination system to facilitate this data flow, among healthcare professionals or family members.

In New Zealand, most medical lab results are electronically transferred directly into practice management systems (PMS) of the GPs and specialists who order the tests. The participant comments in our study indicate that genetic test results are not always included in this system. Because of this, the test results are not delivered to healthcare providers in a timely manner, which delay the data distribution for quality healthcare delivery. A medical testing lab director points out the inability or unwillingness of genetic testing labs to issue reports electronically; and believes that the report should be distributed in the same way as all other lab reports, by electronic (or perhaps paper) transmission to the requesting clinician. On the other hand, it is claimed by genetic testing labs that all results are issued electronically through Delphic reporting system (Sysmex New Zealand Ltd., 2009); yet regional genetic services receive all results on paper. The inconsistency of the data management approach throughout the genetic services delivery system seems to be causing difficulties in data sharing.

In terms of patient's access to test results, the idea of an electronic patient access portal is proposed by both patients and healthcare providers in our interviews. The active involvement of patients in their own health management is believed to have potential for improving health outcomes and cost-effectiveness of medicine. A related topic is the support patients need when dealing with genetic information, including information support and emotional support. Patient support groups are highly

regarded by a patient who has her breasts preventively removed based on BRCA1 positive result and after hyperplasia was found,

> There needs to be support readily available for other women/men who have experienced the same journey... It really lacked that for me. All I wanted was to meet someone my age who lost her breasts and could tell me it would be ok. ... I don't think there is enough focus on the options available to women. I was handed brochures which were 10 years old, not given any support group options, … Most women I talk to say they feel alone until they meet other women who are in the same boat. It is easy for someone to explain your options from a piece of paper but there needs to be detailed information available for these women.

As this patient articulates, information support is needed in order to understand genetic testing and test result, to apply genetics knowledge in their health management so as to improve their health outcome. Moreover, patients have great needs for emotional support, particularly from other patients who are dealing with similar positive genetic test results and are facing similar decision-making.

Regarding family communication about genetic test results, we find that the data distribution firstly depends on the patient's decision whether to spread the word or not. Alternatively, in some Māori families, this decision is made by the elders in the Whānau (extended family). There is no systematic support in place to facilitate this family communication task. One genetic counsellor interviewed reported a case wherein a terminal cancer potentially could have been prevented if the patient had been informed about the risk by their family members. They often did not know that the condition was in the family because they had lost contact with another branch of their family. Our second finding on genetic data dissemination in families is that the task of contacting extended families and telling younger generations is impeded by the inadequate support for information storage and sharing. For instance, patients feel that they are not provided with sufficient educational materials regarding the implications of a test result, including the implications for families and particularly with recommendations for surveillance, management, intervention or cascade testing. This has made the distribution of genetic information a difficult job, especially in extended families, as expressed by a parent of a child patient with cystic fibrosis (CF):

> Most people have a difficult time explaining to extended family the importance of them getting genetically tested for CF. Family are either ignorant/naive (i.e. "it's your problem not ours") or feel that you are just passing blame. … I keep meaning to ask [managers] at the CF Assn [Cystic Fibrosis Association of New Zealand] if there is some information available that we can give to extended family in order for them to understand the genetic side of CF and how easily they can be tested to avoid another baby in the family being born with cystic fibrosis.

This demand for family communication support regarding genetic test results is confirmed by a study in Australia where letters and booklets were found useful as supporting information (Gaff et al., 2005). In addition to these challenges for individuals' distribution of genetic test result, our interviews also identify barriers for aggregated knowledge sharing.

3.3 Difficulties in Knowledge Sharing with Global Knowledge-bases

Genetic information and the associated clinical data collected in clinical genetic services are of high quality and of high relevance to the genetics research community (Gokhale et al., 2004). This presents an opportunity as well as challenge for data dissemination support from routine medical genetic testing up into literature and databases. Some of our participants in labs, genetic service offices, and clinics occasionally conduct research projects and contribute to journal publications; but this is difficult because they are very busy. Our participants report that there is no way of rewarding or encouraging genetic data sharing (e.g. encouraging online database submissions). From a testing lab's point of view, "If it were easy to do it, and particularly if there was some benefit to us, then we might do it. But at the moment it's neither easy nor much benefit to us." To create a system to reward data sharing may require international consensus in the research community. And IT support is needed to facilitate data submission. The participating lab scientists agreed that an ideal genetic information management system should respect patient privacy, should support lab report distribution to healthcare providers, and should assist automatic data dissemination into public databases.

We found that lab storage of genetic data often uses legacy packages from traditional medical test reporting systems and generic office information systems. For example, our participating lab scientists often keep a list of mutations and their interpretation in Microsoft Word, Access, or Excel. One problem with these methods is the lack of data structure, which leads to the flat files in individual labs and subsequently in public databases. Software applications that are specialised in genetic data analysis, such as Vector NTI (Invitrogen Corporation, 2008), Alamut (Interactive Biosoftware, 2009), and Geneious (Biomatters Ltd., 2005, Drummond, 2007) may also hold some fragments of genetic test data in the labs. But these tools don't offer efficient submission support to international databases either.

4 Opportunity for KM Tools

Building upon the above themes that emerged from interviews, this section triangulates our findings with broader health informatics literature. We focus on identifying potential IT solutions that may improve genetics knowledge management, such as decision support systems, electronic referral systems, electronic health record or personal health record systems, data submission and other knowledge processing tools, ontologies, and knowledge networking tools.

4.1 Decision Support Systems (DSS)

This study highlights the need to develop clinician competencies as genetic service providers. Our findings

suggest an opportunity for DSS developers to answer GPs' requests for risk and prognosis information support, as well as for process support in identifying and managing the patients at risk. A key task in delivering genetic services is the test referral decision-making. DSS may assist the task by providing clinicians with information on clinical utility of the test and probability of disease risk, as well as on prevalence, penetrance, and cost/benefit of the test. Another potential for DSS is to collect the right information for genetic counsellors and genetic testing labs. For instance, the family history validation process often occurs after the first meeting between a patient and a genetic counsellor. This process may start at the point of doctor referral by appropriate prompts from a DSS. In addition, DSS may provide information on the availability of genetic tests in the format of a user manual or "Lab Yellow Pages" (Pagon, 2006).

Applying up-to-date knowledge that links genetic tests to health management, DSS should aim not only to prompt referrals for testing, but also to translate test results into intervention recommendations such as surveillance, treatment, and drug dosage. Therefore, future DSS should incorporate evidence-based clinical guidelines and provide intervention recommendations. To provide recommendations rather than just assessments at the time and location of decision-making is a critical success factor for DSS (Kawamoto et al., 2005). The contribution of DSS algorithms to administrative functions such as prompting clinicians about surveillance schedules should be used in genetic testing in much the same way that they are used in primary care, such as recalling patients for cervical screening every three years. This may profoundly and beneficially change clinical practice by realising the value of medical genetics and pharmacogenetics knowledge. For instance, DSS may facilitate complying with established guidelines, such as warfarin initial dosing (US Food and Drug Administration, 2007b, US Food and Drug Administration, 2007a). The implementation of DSS in supporting this task may elicit a wider uptake of genetic testing technology in health care and deliver better health outcomes. Future DSS should also automatically check knowledge-base updates, inform healthcare providers about new scientific findings, identify patients with the relevant mutations that were previously interpreted as uninformative variants, and suggest further actions such as preventive procedures.

4.2 Electronic Referral Systems (E-Referral)

In addition to including DSS functions as part of an e-Referral, the process of referral itself also requires a systematic approach to better support rapidly growing demand for genetic testing. E-Referrals may support streamlining the test referral, and referral triage, processes, may engender consistent business processes in genetic services, as well as may facilitate the tracking and sharing of test status. The New Zealand Health Information Strategy Advisory Committee (HISAC) envisaged that e-Referrals will include features such as referrals tracking, electronic status reports, acknowledging referral receipts, and generating alerts if service level timeframes are not met (New Zealand

Health Information Strategy Advisory Committee, 2009). Better information sharing and communication support would also help bridge the conventional boundaries of primary, secondary and tertiary care, improve provider-user trust and relationships, and ultimately would enhance the knowledge network in the genetic services delivery system.

4.3 Electronic Health Record (EHR) or Personal Health Record (PHR) Systems

The Human Genetics Society of Australasia (HGSA) guidelines suggest genetic data should be kept separately in the genetic services and away from everyday clinical practice on the basis of the nature of the data and the consequences of misuse (Human Genetics Society of Australasia, 2007). Even though this is not a bad idea, it introduces the need for interoperability and for clearly defined rules regarding data sharing. Taking into account also the need to support family communication, we propose adopting shared EHR such as the United Kingdom (UK) HealthSpace SCR (UK National Health Service, 2009) and the United States (US) My HealtheVet (US Department of Veterans Affairs, 2009), or PHR such as Google Health (Google, 2009) and Microsoft HealthVault (Microsoft Corporation, 2010, Quest Diagnostics Inc., 2009) for electronically storing and sharing genetic test results. These systems feature secure storage and transfer of sensitive medical data, particularly with access control to lab results. Especially, PHR has the potential to reduce medical errors, improve disease management, and reduce overall health care costs by empowering patients as active participants in their own health care and facilitating communication between clinicians and patients (Reti et al., 2009, Markle Foundation, 2003). They are aligned with patient-driven health care models (Swan, 2009), which might be applicable for managing genetic test results, given the nature of genetic information as sensitive (personal and medical) and shared (with implications for family members).

4.4 Future KM Targets

Based on participant comments, medical genetics KM requires efforts to facilitate knowledge processes (e.g., to establish knowledge stocks and to support knowledge flows) and to enhance knowledge networks (e.g., to develop core competencies).

4.4.1 Knowledge Processing

To accelerate the knowledge processes of creation, integration, and reuse, one beginning point is to support data dissemination for research purposes. Clinical genetic services have the potential to create, validate and reuse genetics knowledge through collecting genetic and clinical data for genetics research community. The research use of these data may enhance our understanding of the underlying biological mechanisms of human diseases, thus, may lead to better management of the conditions. A key to realizing information sharing, as pointed out by lab scientists, is to make it easy and beneficial to data submitters. For instance, opportunities present themselves to develop systematic support for

database submission to central databases, locus specific databases, or to research projects including individual studies and coordinated projects such as the Human Variome Project (Human Variome Project, 2007). The data dissemination needs have to be appropriately balanced against individual confidentiality concerns. Therefore, de-identification of genetic data is a key to complying with privacy principles.

KM technologies might also support several other processes in the human genetics knowledge cycle, such as where to find knowledge, how to classify, store, maintain, and use it, how to ensure its quality, and how to motivate people to contribute. The codification of this knowledge may depend on standard data presentation in the domain. One key advance towards standardized documentation is the application of gene ontologies.

4.4.2 Gene Ontologies

Because of the complexity of genetic data, standards are essential for human interpretation of genetic test results and for enabling interoperability between information systems. International standards are applied in testing labs, such as mutation nomenclature by Human Genome Organization (Antonarakis and Nomenclature Working Group, 1998, Dunnen and Antonarakis, 2000). However, interpretation of test results is not systematically coded in lab reports. A potentially useful tool for this problem is the well defined gene ontologies from the bio-informatics domain. By applying ontology-controlled terms and concepts, genetic sequence (variant) data would be represented within structure and semantics. Bio-ontologies will serve as: controlled vocabularies for annotation, the describing schema and the content of schema, domain maps, query mechanisms, resolution of semantic heterogeneity, and text analysis (Stevens et al., 2003). The success of the gene ontology (GO) project (Smith et al., 2007) in the bioinformatics research context is yet to be transferred into healthcare settings, especially for annotating the clinical significance of genetic variants. Meanwhile, health level seven (HL7) has organised a Clinical Genomics Special Interest Group, and is developing genotype models and a Family History Model in the interests of interoperability (Shabo, 2007). Further development of clinical genetics terminology using the systematized nomenclature of medicine (SNOMED) (International Health Terminology Standards Development Organisation, 2008) and the logical observation identifiers names and codes (LOINC) (Cimino et al., 2009, Bodenreider, 2008) may also help establish structured genetic test results as transferable health data.

4.4.3 Knowledge Networking

The issues that emerged in our study provide indications of how to enhance the knowledge network among all stakeholders in the genetic services system. First of all, the New Zealand genetic services, as a system, demonstrates ineffective communication and a lack of cooperation among health professionals, which is consistent with a UK report (Eeles et al., 2007). E-Referrals have the potential to facilitate cooperation across sectors, e.g., by sharing test status information, and

therefore to enhance the network of relationships among genetic services stakeholders. A second target in knowledge networking is to prepare healthcare professionals, especially primary and secondary providers, for genetic services delivery. As reported in the UK, a continued lack of awareness of cancer genetics is found amongst primary care practitioners (Iredale et al., 2007). Service protocols and guidelines as well as tools such as DSS may assist care givers in disease risk assessment, genetic testing referral decision-making, and patient management.

From the patient perspective, systematic assistance is needed for enhancing their communication with care givers and family members, e.g., through PHR. Patient contact with other patients in similar circumstances may be improved through online patient support groups or other social networking tools. Consumer health informatics technologies may be useful, including patient support groups such as PinkHope (Barter, 2010), and social networking websites such as PatientsLikeMe (PatientsLikeMe, 2010). With enhanced knowledge networks that involve patients and families, better health outcomes can be expected with more informed intervention choices such as preventive and personalised medicine.

4.4.4 Genetic Services KM Framework

As a summary for the KM targets and promising IT tools based on the issues addressed in this study, we propose a KM framework for the clinical genetic services delivery system, as shown in Figure 1.
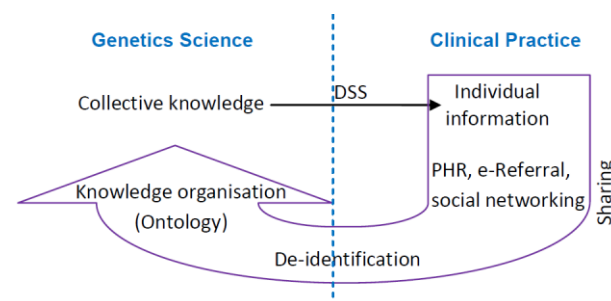
Figure 1: KM framework for the clinical genetic services system

Knowledge flow from genetics science to clinical practice will be supported by tools such as DSS for clinicians during key tasks in genetic service delivery. Feeding back to the body of the human genetic variation knowledge, data sharing will be facilitated by PHR among family members, by e-Referrals across health sectors for test status updates, by social networking tools such as online patient support groups, and by gene ontologies and submission support towards knowledge-bases with appropriate de-identification approach. Better health outcomes are expected from improved capacity of the genetic services system through technology implementation.

5 Discussion

The goal of this study was to examine the use of information relating to genetic services, identifying gaps, and exploring differences between "usual" health

information management and medical genetic information management so as to provide groundwork for refining KM processes to meet the needs of people using and working in genetic services delivery. Our findings have highlighted challenges surrounding medical genetic data sharing and usage, allowing us to suggest IT opportunities for supporting human genetic variation knowledge management. Promising tools include DSS, e-Referrals, and EHR or PHR. Implementing these health informatics technologies, as well as KM tools (e.g., data submission and other knowledge processing tools, ontology, and knowledge networking tools), would address the knowledge processes in the domain and potentially enhance knowledge networks.

A closely related investigation was the 2003 GP survey in New Zealand regarding their knowledge and practice on genetic testing (White and McLeod, 2003, Morgan et al., 2004). This national survey found that GPs felt that they lacked experience and knowledge of genetic testing, especially on the appropriate terminology and procedures. Our interviews have found little change regarding GP involvement in the genetic services delivery system. It appeared that clinicians, including GPs, need help to identify the utility of genetic tests. A 2009 study in the US highlighted the promotion of the inclusion of genetic information in the EHR (Sethi and Theodos, 2009). Our study concludes that IT might have great potential in the domain, such as DSS, e-Referrals, EHR/PHR, and KM tools. In particular, national implementation of DSS and e-Referrals is part of the current New Zealand national health IT plan (IT Health Board, 2010).

The main limitation of this study is that it is based upon personal perspectives from individual experience which might not represent accurately the entire New Zealand genetic services system. Strategic planning and framework development are needed to enable more widespread participation in the active use of genetic information among healthcare professionals, and to achieve a degree of uniformity and interoperability among nations. Establishment of such strategies and frameworks may require more research into the technical, organisational, and social issues that will open up public debate on the topics.

Future actions on the part of particular stakeholder groups include:

- *Clinicians* should require information and systematic support for the processes of genetic services, including electronic decision support and effective electronic referral systems, as well as clarification of ethical and legal framework.

- *Health IT innovators* should develop and implement quality systematic solutions to support key tasks in genetic service delivery.

- *Policy makers* should lead wide debates on the legal and ethical issues surrounding genetic information in order to develop appropriate frameworks that will facilitate better use of genetic data to improve health care.

Furthermore, patients and their families, of course, have the most at stake. They – as citizens generally, and through specific groups, such as cancer societies – should insist on action on the three fronts above, as well as with respect to information, information management technology and support groups for their own direct use.

6 Conclusion

Taking a grounded theory approach, we explored the domain of human genetic variation knowledge management in the New Zealand clinical genetic services system. Clinicians' lack of knowledge of genetic testing and the lack of support for test result dissemination present barriers for unleashing the potential of genetics knowledge in health care. To overcome these challenges, we recommended a few KM technologies, including DSS, e-Referrals, EHR/PHR systems, data submission and other knowledge processing tools, ontologies, and knowledge networking tools. These will facilitate genetics knowledge processing by supporting the flow of genetic data such that it is appropriately structured and interpretable to all stakeholders. Moreover, knowledge networks among clinicians, genetic counsellors and lab scientists, as well as with patients and family members, will be enhanced through better communication support and more available information.

7 Acknowledgements

8 References

ALAVI, M. & LEIDNER, D. E. (2001) Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *MIS Quarterly,* 25**,** 107-136.

ANTONARAKIS, S. A. & NOMENCLATURE WORKING GROUP (1998) Recommendations for a Nomenclature System for Human Gene Mutations. *Human Mutation,* 11**,** 1-3.

ARGOTE, L., MCEVILY, B. & REAGANS, R. (2003) Managing Knowledge in Organizations: An Integrative Framework and Review of Emerging Themes. *Management Science,* 49**,** 571-582.

BARTER, K. (2010) Pink Hope: Every journey is different but we all share the same hope.

BELL, J. (1998) The new genetics: The new genetics in clinical practice. *BMJ,* 316**,** 618-620.

BHANDARI, M., GARG, R., GLASSMAN, R., MA, P. C. & ZEMMEL, R. W. (2000) A Genetic Revolution In Healthcare. *Healthcare Quarterly,* 3**,** 58-62.

BIOMATTERS LTD. (2005) Geneious: Research in a flash.

BODENREIDER, O. (2008) Biomedical ontologies in action: role in knowledge management, data integration and decision support. *Yearb Med Inform***,** 67-79.

CARLSSON, S., ELSAWY, O., ERIKSSON, I. & RAVEN, A. (1996) Gaining Competitive Advantage Through Shared Knowledge Creation: In Search of a New Design Theory for Strategic Information Systems. *Proceedings of the 4th European Conference on Information Systems.* Lisbon, Portugal.

CIMINO, J. J., HAYAMIZU, T. F., BODENREIDER, O., DAVIS, B., STAFFORD, G. A. & RINGWALD, M. (2009) The caBIG terminology review process. *Journal of Biomedical Informatics,* 42**,** 571-580.

CRESWELL, J. W. (2003) *Research design : qualitative, quantitative, and mixed methods approaches* Thousand Oaks, Calif., Sage Publications.

DAVENPORT, T. H. & PRUSAK, L. (1997) *Working Knowledge: How Organizations Manage What They Know,* Cambridge, MA, Harvard Business School Press.

DRUMMOND, A. (2007) Publications and Software.

DUNNEN, J. T. D. & ANTONARAKIS, S. E. (2000) Mutation nomenclature extensions and suggestions to describe complex mutations: A discussion. *Human Mutation,* 15**,** 7-12.

EELES, R., PURLAND, G., MAHER, J. & EVANS, D. G. (2007) Delivering cancer genetics services-new ways of working. *Familial Cancer,* 6**,** 163.

ENGSTROM, J. L., SEFTON, M. G. S., MATHESON, J. K. & HEALY, K. M. (2005) Genetic competencies essential for health care professionals in primary care. *Journal of Midwifery & Women's Health,* 50**,** 177-183.

FEERO, W. G., BIGLEY, M. B. & BRINNER, K. M. (2008) New Standards and Enhanced Utility for Family Health History Information in the Electronic Health Record: An Update from the American Health Information Community's Family Health History Multi-Stakeholder Workgroup. *Journal of the American Medical Informatics Association,* 15**,** 723-728.

GAFF, C., COLLINS, V., SYMES, T. & HALLIDAY, J. (2005) Facilitating Family Communication About Predictive Genetic Testing: Probands' Perceptions. *Journal of Genetic Counseling,* 14**,** 133-140.

GLASER, B. G. & STRAUSS, A. L. (1967) *The discovery of grounded theory; strategies for qualitative research* Chicago, Aldine Pub. Co.

GLASNER, P. (2002) Beyond the genome: reconstituting the new genetics. *New Genetics and Society,* 21**,** 267 - 277.

GOKHALE, D. A., DEVEREAU, A. D. & TAYLOR, G. R. (2004) A new online mutation repository for UK diagnostic laboratories.

GOOGLE (2009) About Google Health. Google.

GU, Y., WARREN, J. & DAY, K. (2010) Helping Clinicians Identify the Clinical Utility of Genetic Tests. *the HISA Health Informatics Conference 2010.* Melbourne, Australia.

GU, Y., WARREN, J. & DAY, K. (2011) Unleashing the Power of Human Genetic Variation Knowledge: New Zealand Stakeholder Perspectives. Paper accepted by Genetics in Medicine. 2011:13(1). DOI: 10.1097/GIM.0b013e3181f9648a.

HUMAN GENETICS SOCIETY OF AUSTRALASIA (2007) Position Statement and Guidance: The Protection of Human Genetic Information.

HUMAN VARIOME PROJECT (2007) What is the Human Variome Project? *Nat Genet,* 39**,** 423-423.

INTERACTIVE BIOSOFTWARE (2009) Alamut - Mutation Interpretation Software.

INTERNATIONAL HEALTH TERMINOLOGY STANDARDS DEVELOPMENT ORGANISATION (2008) International Health Terminology Standards Development Organisation (IHTSDO): Our Standards - SNOMED CT. International Health Terminology Standards Development Organisation (IHTSDO).

INVITROGEN CORPORATION (2008) Vector NTI Advance™ - Sequence analysis and data management software for PCs.

IREDALE, R., ELWYN, G., EDWARDS, A. & GRAY, J. (2007) Attitudes of genetic clinicians in Wales to the future development of cancer genetics services. *Journal of Evaluation in Clinical Practice,* 13**,** 86-89.

IT HEALTH BOARD (2010) National Health IT Plan. http://www.ithealthboard.health.nz/sites/all/files/National%20Health%20IT%20Plan%20v11_1.pdf.

IT PERSPECTIVES INC. (2005) NHS Genetic Service Information Systems Output Based Specification.

KÄÄRIÄINEN, H., RANTANEN, E. & HIETALA, M. (2006) Summary of the guidelines for genetic counselling.

KAWAMOTO, K., HOULIHAN, C. A., BALAS, E. A. & LOBACH, D. F. (2005) Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ,* 330**,** 765.

MARKLE FOUNDATION (2003) Connecting for health in the information age: the personal health working group final report. Sec III. Why: Making the case for the personal health record. New York.

MCGUIRE, A. L., FISHER, R., CUSENZA, P., HUDSON, K., ROTHSTEIN, M. A., MCGRAW, D., MATTESON, S., GLASER, J. & HENLEY, D. E. (2008) Confidentiality, privacy, and security of genetic and genomic test information in electronic health records: points to consider. *Genetics in Medicine,* 10**,** 495-499.

MICROSOFT CORPORATION (2010) Welcome to Microsoft HealthVault.

MORGAN, S., MCLEOD, D., KIDD, A. & LANGFORD, B. (2004) Genetic testing in New Zealand: the role of the general practitioner. *Journal of the New Zealand Medical Association,* 117.

NEW ZEALAND HEALTH INFORMATION STRATEGY ADVISORY COMMITTEE (2009) eReferrals.

NONAKA, I. (1994) A Dynamic Theory of Organizational Knowledge Creation. *Organization Science,* 5**,** 14-37.

PAGON, R. A. (2006) GeneTests: an online genetic information resource for health care providers. *J Med Libr Assoc,* 94**,** 343-348.

PATIENTSLIKEME (2010) Find Patients Just Like You.

PERSONALIZED MEDICINE COALITION (2006) The Case for Personalized Medicine.

QUEST DIAGNOSTICS INC. (2009) Quest Diagnostics to Connect to Microsoft HealthVault and Empower Patients and Physicians to Share Diagnostic Laboratory Test Results Online.

RETI, S. R., FELDMAN, H. J. & SAFRAN, C. (2009) Governance for Personal Health Records. *Journal of the American Medical Informatics Association,* 16**,** 14-17.

RHINE, J. (2008) The Grounded Theory Institute: The official site of Dr. Barney Glaser and Classic Grounded Theory. .

SAMBAMURTHY, V., BHARADWAJ, A. & GROVER, V. (2003) Shaping Agility Through Digital Options: Reconceptualizing the Role of Information Technology in Firms. *MIS Quarterly,* 27**,** 237-263.

SCHULTZE, U. & LEIDNER, D. (2002) Studying knowledge management in information systems research: Discourses and theoretical assumptions. *MIS Quarterly,* 26**,** 213-242.

SETHI, P. & THEODOS, K. (2009) Translational bioinformatics and healthcare informatics: computational and ethical challenges. *Perspect Health Inf Manag,* 6**,** 1h.

SHABO, A. (2007) HL7 Clinical Genomics Genotype Models.

SMITH, B., ASHBURNER, M., ROSSE, C., BARD, J., BUG, W., CEUSTERS, W., GOLDBERG, L. J., EILBECK, K., IRELAND, A., MUNGALL, C. J., LEONTIS, N., ROCCA-SERRA, P., RUTTENBERG, A., SANSONE, S.-A., SCHEUERMANN, R. H., SHAH, N., WHETZEL, P. L. & LEWIS, S. (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology,* 25**,** 1251-1255.

STANKOSKY, M. A. (2005) Advances in Knowledge Management: University Research Toward an Academic Discipline. IN STANKOSKY, M. A. (Ed.) *Creating the Discipline of Knowledge Management: The Latest in University Research.*

Burlington, MA., Elsevier ButterworhHeinemann.

STEVENS, R., WROE, C., LORD, P. & GOBLE, C. (2003) Ontologies in bioinformatics. IN STAAB, S. & STUDER, R. (Eds.) *Handbook on Ontologies.* Springer.

STRAUSS, A. L. & CORBIN, J. M. (1998) *Basics of qualitative research : techniques and procedures for developing grounded theory,* Thousand Oaks, Calif., Sage Publications.

SWAN, M. (2009) Emerging Patient-Driven Health Care Models: An Examination of Health Social Networks, Consumer Personalized Medicine and Quantified Self-Tracking. *International Journal of Environmental Research and Public Health,* 6**,** 492-525.

SYSMEX NEW ZEALAND LTD. (2009) Delphic - Streamlining your laboratory.

UK NATIONAL HEALTH SERVICE (2009) NHS HealthSpace - Your Summary Care Record. The National Health Service (NHS).

US DEPARTMENT OF VETERANS AFFAIRS (2009) My HealtheVet

US FOOD AND DRUG ADMINISTRATION (2007a) COUMADIN® TABLETS (Warfarin Sodium Tablets, USP) Crystalline; COUMADIN® FOR INJECTION (Warfarin Sodium for Injection, USP).

US FOOD AND DRUG ADMINISTRATION (2007b) November 14-15, 2005 Clinical Pharmacology Subcommittee.

WHITE, S. & MCLEOD, D. (2003) Genetic Testing: A Survey of New Zealand General Practitioner's Knowledge and Current Practice. Wellington, National Health Committee.

ZIMMERN, R. L. (1999) The human genome project: a false dawn? *BMJ,* 319**,** 1282-.

Chronic Disease Management: a Business Intelligence Perspective

Leelani Kumari Wickramasinghe[1]   Damminda Alahakoon[2]   Michael Georgeff[1]

Peter Schattner[1]   Daswin De Silva[2]   Oshadi Alahakoon[2]   Akuh Adaji[1]

Kay Jones[1]   Leon Piterman AM[1]

[1] Department of General Practice
Monash University,
Melbourne, Australia.
Email: Leelani.Wickramasinghe@monash.edu

[2] Clayton School of Information Technology
Monash University,
Melbourne, Australia.
Email: Damminda.Alahakoon@monash.edu

Abstract

Chronic disease management is one of the main areas in healthcare that Health Knowledge Management (HKM) can provide beneficial outcomes. Information Communication Technology (ICT) enabled Chronic disease management network (cdmNet) delivers comprehensive chronic disease management solutions by incorporating all key processes of the Chronic Care Model (CCM) developed by Wagner and his group. Through its roll-outs, cdmNet has accumulated detailed data about the chronic disease management process. This paper presents a new Business Intelligence (BI) module developed to analyse, visualise and extract knowledge from the cdmNet data. The aim of the BI module is to facilitate the short-term and long-term decision making and improve understanding of collaborative care models, policy and economic models underlying chronic disease management. The paper contains preliminary results obtained from applying the BI module to the cdmNet data.

Keywords: Chronic disease management, data mining, business intelligence

1 Introduction

Healthcare Knowledge Management (HKM) is defined as the systematic creation, modeling, sharing, operationalization and translation of healthcare knowledge with the emphasis to improve the quality of patient care [Sibte and Abidi, 2008]. The aim of HKM is to provide high quality, well-informed and cost-effective patient care decisions to healthcare stakeholders such as government, healthcare professionals and even patients themselves. Due to the increasing number of patients with chronic disease and the associated medical care costs [Anderson and Wilson, 2006, Anderson and Johnson, 2004], chronic disease management is one among many other areas of healthcare, that HKM can provide beneficial outcomes.

The Chronic Care Model (CCM) developed by Wagner and his group [Wagner et al., 2001a] em-phasises collaboration among care providers and the patient in creating and maintaining a care plan for patients with chronic disease. A recent Australian initiative, chronic disease management network (cdmNet) [Georgeff et al., 2010, Georgeff and Hilton, 2010] is an ICT implementation to put the CCM into practice. cdmNet focuses on the processes which are key to the CCM by changing chronic disease management domain from point-to-point, episodic, referral to a continuous, collaborative, networked model. Through its already rolled-out phases and future roll-outs, cdmNet, for the first time has accumulated detailed data about the process of chronic care management. The knowledge concealed within the raw data have the potential to provide an evidence base for the CCM and could be the key to informed health policy for patients with chronic disease.

Business Intelligence (BI) refers to computer-based techniques to spot, drill down (for detailed information), roll up (for abstract information) and analyse business data [Larson, 2008]. BI technologies are incorporated with functions such as reporting, on-line analytical processing, analytics and data mining to provide historical, current and predictive views of business operations.

The aim of this paper is to demonstrate the application of BI techniques to systematically create, model, share and translate data generated from chronic disease management process which enables data driven decision making. Application of BI techniques to the cdmNet data is threefold: (1) to interpret prevailing chronic disease management process from stakeholders' perspectives; (2) to extract interesting patterns hidden within data; and (3) to provide a basis for validating the Chronic Care Model. All these are key for: (1) short-term and long-term decision making; and (2) improving understanding of collaborative-care models, policy, and economic models underlying chronic disease management. The cdmNet-BI module presented in the paper aims to fulfil the aforementioned three criteria, thus providing guidance to improve the quality of chronic disease patient care.

The paper is organised as follows: Section 2 provides background material on chronic disease management and the cdmNet system. The cdmNet Business Intelligence (cdmNet-BI) module is described in Section 3. Section 4 presents the preliminary results obtained by applying this BI module to the cdmNet data. Concluding remarks and future work are discussed in Section 5.

2 Background

The basis of this paper is chronic disease management and the cdmNet system which accumulates chronic disease process related data. Brief introductions to chronic disease management and cdmNet system are provided in Sections 2.2 and 2.1 respectively.

2.1 Chronic Disease Management

Chronic diseases such as type 2 diabetes, asthma, and arthritis generally cannot be cured completely. Hence they persist through patients' lifetime and require adequate management [2007]. In Australia, the management of a patient with a chronic disease, tends to emphasis on following the best-practice clinical guidelines. Once an individual is identified with a chronic condition that requires ongoing management, a General Practitioner (GP) may tend to create a General Practice Management plan (GPMP) (a care plan in the context of this paper) from their assessment of the patient's underlying conditions and care objectives. Care plans include goals, strategies and specific tasks. These will include services by other care providers (e.g., diabetic education, podiatry service), clinical interventions by the GP (e.g., tests such as lipid tests ordered) and any medication required for managing the chronic condition. Once the GPMP is created, the GP identifies possible healthcare providers who can provide the services listed in GPMP and creates a formal document called a team care arrangement (TCA). TCA includes the details of healthcare providers who provide each service identified in GPMP. GPMPs and TCAs are known as Medicare Benefits Schedule (MBS) items [2009]. Over the course of the GPMP/TCA lifetime the GP reviews the execution of the plan, follows up whether the patient is attending appointments (either directly or through a carer), whether key tests or treatments have been performed, and monitors patient's condition for changes.

Prior to cdmNet and in the regions in which cdmNet is not yet rolled-out, GPs use chronic disease management care plan templates available in GPs' desktop application for care plan creation. Template based chronic disease management is not efficient and does not comply with the CCM due to following limitations:

- Templates provide a guideline, requiring some manual personalisation that demands GP's time;

- Lack of care coordination; and

- Either limited or no support for patients in adhering to care plans ensuring appointments are made, visits are attended and medications are renewed.

cdmNet provides solution to these issues by [Georgeff and Hilton, 2010]:

- Creating best practice, personalised care plans;

- Distributing care plans to the patients' care team and to the patients;

- Monitoring continuously the care plan, medication renewals, and appointments;

- Ensuring timely follow up;

- Facilitating collaboration by sharing the health record, care plan, and progress against the care plan among the care team and with the patient; and

- Supporting patient self-management by sending alerts, reminders, and notifications.

cdmNet was trialled in the Barwon South Western Region (BSWR) of Victoria and the Eastern Goldfields Region (EGR) of Western Australia. These trials involved 97 GPs, 268 other healthcare providers (including practice nurses), and 733 patients with diabetes. cdmNet is currently being rolling out, covering a population of over 1.2 million, in Melbourne metro region and regional Victoria, Queensland, Tasmania and Western Australia.

2.2 Chronic Disease Management Network (cdmNet)

The aim of cdmNet is to reduce the problems associated with template based chronic disease management by implementing most of the elements of the CCM. cdmNet identifies processes as the key to the CCM and breaksdown the CCM processes to stages as follows [Georgeff and Hilton, 2010]:

- Planning: identifies chronic disease subpopulation and creates best-practice, personalised care plans;

- Collaboration: identify healthcare providers to be part of the care plan, develop agreements and distribute care plans among the care providers; and

- Monitoring: track compliance with care plan by care team and patient in real-time, monitor key patient health parameters (e.g., blood glucose) and support adherence by sending alerts, reminders and notifications.

Through these processes, cdmNet accumulates data on:

- Patients: demographics (e.g., gender, age, marital status), lifestyle (e.g., drinking status, smoking status), measurements (e.g., blood pressure, body weight), and medications (e.g., name, strength, dose);

- Items on care plans: service, goal, target, care provider, frequency;

- Appointments with providers: With whom, when, service, status;

- Information about MBS items: type, status;

- Use of cdmNet adherence support services: type (e.g., email, sms), recipient, when; and

- cdmNet web page accesses: accessor, when, which page

The aim of cdmNet-BI module is to harvest tacit knowledge concealed within this detailed data about the process of chronic care management for decision support. Several key performance indicators, such as the use of cdmNet for care planning by GPs, have been established to ascertain the performance of cdmNet.

3 cdmNet Business Intelligence Module (cdmNet-BI Module)

The purpose of the cdmNet-BI module is to convert above mentioned transactional data gathered in and/or generated from cdmNet operational processes to knowledge. It consists of three sub-modules: (1)

pre-processing, (2) dashboard, and (3) data mining as shown in Figure 1. The pre-processing sub module, as described in Section 3.1 converts transactional data to a format that can be used in dashboard and data mining sub modules. The dashboard and the data mining sub modules aim to provide solutions for closed and open questions related to chronic disease management process. Closed questions are answered by distinct values without any uncertainty from the cdmNet data. Solutions to open questions contain certain degree of uncertainty and may have many possible answers.
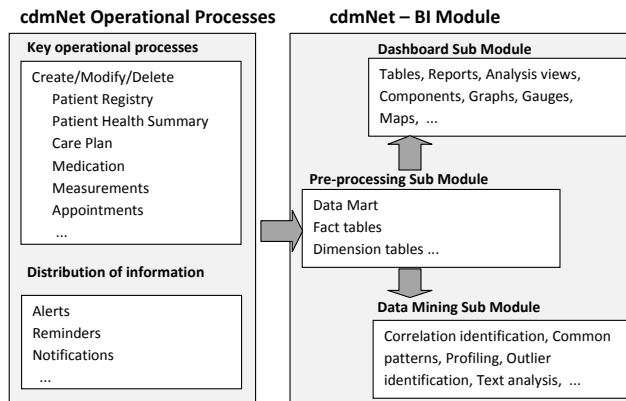

Figure 1: cdmNet-BI module.

3.1 Pre-processing Sub Module

The cdmNet data generated by day-to-day transactions, as any other business data are stored in databases which adopt database normalization rules. When analysing the data, speed of access is the main criteria to be considered. In addition, depending on the analyses to be performed, it may not require all transactional data. Incorporating these requirements, the pre-processing sub module performs two types of functions:

1. Identifies data required for analysis; and

2. Converts data identified in Item (1) above to gain speed of access to suit analysis purposes.

After identifying the data required for analysis, the pre-processing sub module creates a data mart to contain only the required data and populates it using extract, transform, and load (ETL) process [Larson, 2008]. Data mart is a denormalised (repeated) relational database designed for speed of access [Larson, 2008]. When data are denormalised, analysis and reporting can be performed using only a few table joins thus increasing the speed.

A data mart structure is based on four elements: measures, dimensions, attributes and hierarchies [Larson, 2008]. A measure is a numeric quantity that represents some aspect of the operational system. Number of care plans generated, number of over weight diabetes patients, and number of users who access cdmNet web pages are few examples of measures in this application. The measures are stored in tables called fact tables; for example CareplansFact, PatientWeightFact, and PageAccessesFact.

Dimensions are the perspectives in analysing data. For example, often stakeholders are not interested in the total number of care plans generated by the system. Instead they prefer to slice and dice this total into its constituent parts. For example, commonwealth government may be interested in comparing the performances of each state while a particular state government is interested in comparing different general practice divisions within the state. These contributing components such as state and general practice division are known as dimensions.

Attributes refer to any additional information stored about dimension members. For example, general practice division name is an attribute of the general practice dimension. Usually, a dimension is part of a large structure known as a hierarchy. For example, state and division may have their own hierarchies. In summary, all data mart components are created and populated by the pre-processing sub module.

3.2 Dashboard Sub Module

The Dashboard sub module is designed to provide two functionalities:

- to provide solutions to closed questions; and

- to present solutions in line with decision makers' objectives.

Few examples of chronic disease closed questions are:

1. For each month how many care plans are created in a particular state (say Victoria) by a particular GP (say Dr. John Smith) for a given time period (say from 2010-01-01 to 2010-06-30)?

2. Which state has generated the most number of care plans?

3. How many chronic disease patients are listed within a given GP and what are their demographics?

4. Compare the number of care plans created and reviewed by a given GP with state, division and organisation averages; and

5. Compare care planning performances between and among states, divisions and GPs.

As an end result, the dashboard sub module represent all key performance indicators in respect to stakeholders. Easy access and visualization are the main considerations of this sub module. Fulfilling these requirements, the sub module creates portals for different stakeholders. Portals include tables, reports, analysis views, components (radio, select, check), graphics (bar charts, pie charts), gauges (dial charts, traffic light) and maps. All these user interface components visually illustrate key performance indicators.

3.3 Data Mining Sub Module

As any other real world system, chronic disease management process is concerned with open questions, which may have multiple possible solutions with varying degrees of certainty. Few examples of such open questions are:

1. What are the measures that determine health outcomes in chronic disease management? Possible measures can be:

  - changes in patient measurements such as body weight, blood pressure; and/or

  - changes in number of hospitalisations or work force participation;

2. Which variables determine the health outcomes identified in Item (1) above? Is it

- the chronic disease?
- number and type of service in care plan?
- patient's adherence to the care plan?
- regular reviews and follow-ups by GPs?
- patient's interaction with cdmNet web site?

Hence, the aim of the data mining sub module is twofold: (1) to provide possible solutions for open questions; and (2) to extract interesting patterns from data. The outcomes of the latter may in turn provide guidance to the former aim.

Data mining is defined as non-trivial discovery of novel, valid, comprehensible and potentially useful patterns from data [Fayyad et al., 1996]. This sub module aims to use both directed and undirected knowledge discovery techniques [Berry and Linoff, 2004] to provide functionalities such as clustering, classification, prediction, association mining, and text mining [Smith, 2000].

In directed knowledge discovery process, there is a goal which directs the knowledge discovery process. For example, we can select certain variables, such as patients' demographics (gender, age, weight) and lifestyle (smoking status, drinking status) as inputs and blood glucose level as output. A model can be trained to learn the relationship between inputs and output using a subset of the cdmNet data. The rest of the data can be used to validate the knowledge discovered. Once the model is trained to a certain confidence level, it can be used to predict blood glucose level of a new patient when demographics and lifestyle are known.

Undirected knowledge discovery techniques are used to investigate the underlying structure of data or for pattern recognition. These techniques neither distinguish between variables and outcomes nor contain a specific goal. All data are considered as parameters and the technique determines relationships that may exist among the parameters. Currently, the data mining sub module uses clustering as the undirected knowledge discovery technique. Clustering algorithms classify data into groups based on the natural structure of data without obtaining additional information about data from a data analyst [Smith, 2000]. Specifically, self-organizing map (SOM) [Kohonen, 1982a,b] is used as a clustering technique as SOM has the ability to group data patterns based upon the natural associations found in data [Smith, 2000]. In addition, SOM provides cluster visualisation by mapping high dimensional input space (all parameters) to low dimensional feature space. Feature space only contains variables that significantly contributed towards cluster separation.

4 Application of the cdmNet-BI Module to cdmNet Data

After evaluating a number of BI tools, ETL functionalities of the pre-processing sub module and the dashboard sub module are built using Pentaho BI suite community edition [2008]. Pentaho is a free, open source and Java based BI tool. The cdmNet system is built using Java technologies. Thus Pentaho makes it convenient to integrate the BI module with the cdmNet system. Data mart is built using PostgreSQL [1996] database management system as of the cdmNet system. Currently, for clustering purposes the data mining sub module uses Viscovery SOMine [2010] which is reviewed as a commercial standard tool for the SOM algorithm.

4.1 Results from the Pre-Processing Sub Module

This section describes the dimensions that can be used to represent the measure, number of care plans. We identified 7 dimensions: state, postcode, division, practice, GP, time and item to represent the measure (fact), 'number of care plans'. These dimensions provide single and combinations of perspectives to the given measure. An architecture called star schema is generally used for the design of a data mart [Kimball et al., 2002]. A star schema that represents this fact table and its dimensions is shown in Figure 2.
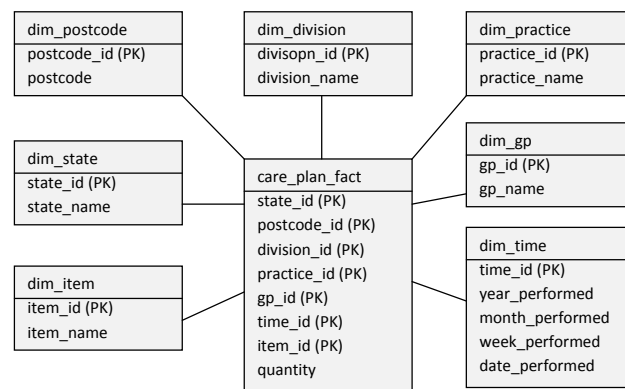

Figure 2: Star schema with dimension and fact tables.

A data mart using PostgreSQL was created to reflect the fact and dimension tables and populated with the cdmNet data.

4.2 Results from the Dashboard Sub Module

This section illustrates several screens developed using the dashboard sub module. An analysis view with drill down (for detailed information) and roll up (for abstract information) capabilities is shown in Figure 3. It is a visual representation of all dimensions and the fact described in Section 3.1. Its navigational capabilities provide stakeholders with the ability to investigate along one or many dimensions depending on their requirements.

A portal developed for the GPs is shown in Figure 4. The table at the top left hand corner shows cumulative monthly revenue and the corresponding line chart is shown at the top right hand corner. The pie chart at the bottom right hand corner shows the percentage of care plan items created by the GP during the current month. The bar chart at the bottom left hand corner compares the GP's performance against the average of all GPs, average of top GPs (who have created more than a specific number of care plans) and the national average.

4.3 Results from the Data Mining Sub Module

Preliminary results obtained from the data mining sub module to identify the factors that contribute to diabetes chronic condition is described in this section. HbA1c or haemoglobin A1C is a metabolic measure used to diagnose diabetes. HbA1c is a molecule created in red blood cells when glucose sticks to them. A normal non-diabetic HbA1c is 3.5-5.5%. With diabetes, about 6.5% is considered normal. Achieving HbA1c $< 7\%$ is considered as a diabetes management goal [Wagner et al., 2001b].

State	Postcode	Division	Practice	GP	Time	Item	Measures Quantity
⊟ All States	⊞ All Postcodes	⊞ All Divisions	⊞ All Practices	⊞ All GPs	⊞ All Times	⊞ All Items	4,031
SA	⊞ All Postcodes	⊞ All Divisions	⊞ All Practices	⊞ All GPs	⊞ All Times	⊞ All Items	149
VIC	⊞ All Postcodes	⊟ All Divisions	⊞ All Practices	⊞ All GPs	⊞ All Times	⊞ All Items	2,539
		Dandenong Casey	⊞ All Practices	⊞ All GPs	⊞ All Times	⊞ All Items	4
		General Practitioners Association of Geelong	⊞ All Practices	⊞ All GPs	⊞ All Times	⊞ All Items	2,400
		Monash Division of General Practice	⊟ All Practices	⊞ All GPs	⊞ All Times	⊞ All Items	103
			East Bentleigh Medical Group	⊞ All GPs	⊟ All Times	⊞ All Items	103
					⊞ 2008	⊞ All Items	4
					⊞ 2009	⊞ All Items	45
					⊟ 2010	⊞ All Items	54
					⊞ 2010 January	⊟ All Items	11
						Allied Health Referral Form	2
						GPMP	3
						GPMP Review	2
						HMR	2
						TCA	2
					⊞ 2010 February	⊞ All Items	10
					⊞ 2010 March	⊞ All Items	3
					⊞ 2010 April	⊞ All Items	3
					⊞ 2010 May	⊞ All Items	1
					⊞ 2010 June	⊞ All Items	14

Figure 3: Analysis view with drill down and roll up capabilities.
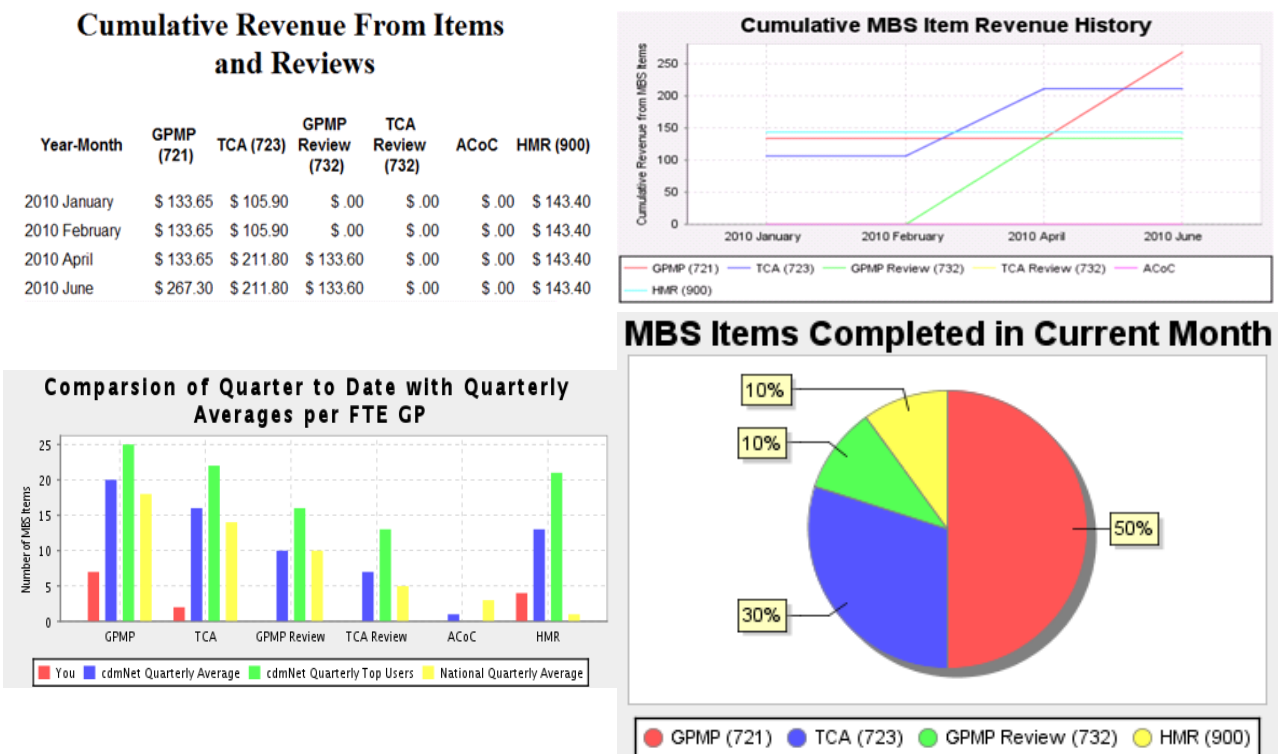

Figure 4: A screen from the GP portal. The table and the line chart show the cumulative monthly revenue of the GP from care plan items for a given time period. The pie chart shows the percentage of care plan items created by the GP during the current month. The bar chart compares the GP's performance against the average of all GPs, average of top GPs (who have created more than a specific number of care plans) and the national average.

Chronic diseases such as diabetes can be directly related to patients' lifestyle and demographics. This experiment investigates the impact of patients' lifestyle and demographics on HbA1c. It was carried out based on three categories of parameters: (1) demographics; (2) lifestyle; and (3) metabolic measures. The parameters considered for each category include:

1. Demographics: gender, age, marital status;

2. Lifestyle: drinking status, smoking status; and

3. Metabolic measure: HbA1c.

To determine the impact, for each patient, HbA1c measured over a time period is considered. That is, the experiment includes patients whose HbA1c level is measured more than once. In the original data set, there are 657 patients with HbA1c readings, but only 338 have measured it more than once. The total number of readings for a given patient varies between 2 and 42. Therefore, data corresponding to these 338 patients were used. In addition to the parameters discussed in Section 4.3, this experiment includes a number of parameters which describe statistical measures of HbA1c and trend in consecutive HbA1c measurements as follows:

- Parameters relate to statistical measures of HbA1c: average HbA1c, minimum HbA1c, maximum HbA1c, and standard deviation of HbA1c; and

- Parameters relate to trend in consecutive HbA1c measurements: percentages of increased consecutive readings, decreased consecutive readings, and stable consecutive readings

For each patient, trend in consecutive HbA1c readings is calculated as:

$$percentage\ of\ increased\ consecutive\ readings = count(V_{t+1} > V_t)/N * 100$$

where V denote HbA1c reading, t denote time and N denote total number of HbA1c readings of that particular patient. Similar calculations are carried out for percentage decrease and percentage stable.

The experiments were carried out using clustering as the knowledge discovery technique, self-organizing map as the clustering technique and Viscovery SOMine application as the tool to obtain SOM clusters as mentioned in Sections 3.3 and 4. A few preprocessing steps were carried out to encode non-numeric data before applying the Viscovery SOMine application as follows:

- gender: binary data (0 - male, 1 - female);

- drinking status: binary data (0 - non drinker, 1 - drinker);

- smoking status: 0 - non smoker, 0.5 - ex-smoker, 1 - smoker; and

- marital status: 1 of N for single, defacto, married, widowed, divorced and separated;

Parameters that take continuous values are transformed using Sigmoid or Logarithmic transformations to obtain normal distributions.

- Sigmoid transformation: age, average HbA1c, minimum HbA1c, maximum HbA1c, percentage increase, percentage decrease, percentage stable; and

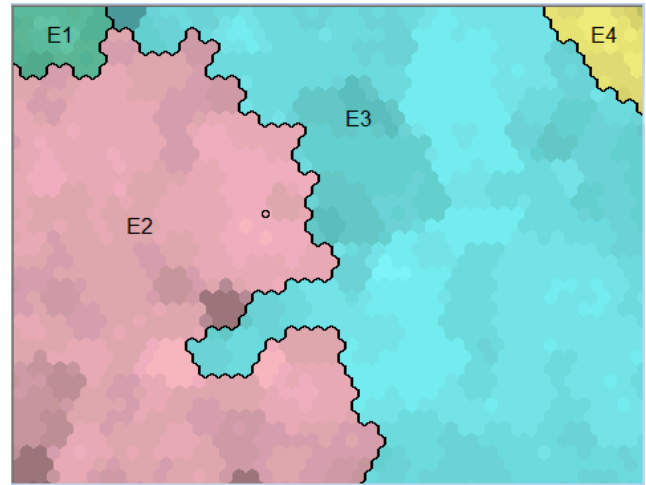- Logarithmic transformation: standard deviation of HbA1c

Figure 5: Visual representation of formed clusters.

A visual representation of clusters generated by the Viscovery SOMine application is shown in Figure 5. It consists of four clusters: E1, E2, E3 and E4. Table 1 contains the average values of each parameter contributing towards the formation of each cluster. Using the values in Table 1, an analysis of each cluster is carried out in Table 2 to determine the reason for the formation of clusters. Such reasoning has the potential of discovering knowledge.

According to Table 2, clusters E1 and E2 include patients with HbA1c > 7%. Patients in E1 has a tendency to increase HbA1c, while patients in E2 have fluctuating readings. The data highlight that while heavy drinking and previous smoking contribute to high HbA1c, whether HbA1c continues to increase or fluctuate is effected by gender and marital status.

Clusters E3 and E4 have patients with HbA1c < 7% and they continue to have stable readings. The data indicate that both moderate or no alcohol and non smoking determines low HbA1c. Age and marital status are the determining factors for seperating the two cluters.

Even though these patterns are not unknown or unforeseen, this technique:

- provides a guideline for stakeholders to identify patients who need more support (education, tests, other health professional visits) to manage their diabetes from patients from patients demographics and lifestyle; and

- identifies hypothesis for further testing, for example married males around 60 years of age who are heavy drinkers and ex-smokers find it hard to control their diabetes.

5 Conclusions and Future Work

This paper presented a Business Intelligence (BI) module developed to analyse, visualise and extract knowledge from the data accrued in the cdmNet system. BI module consists of three sub modules: (1) pre-processing; (2) dashboard; and (3) data mining. The purpose of the pre-processing sub module is to convert the cdmNet data to a form suitable for fast access. The dashboard sub module provides interfaces with drill down, roll up and graphic components. It enables stakeholders to navigate through information to understand the prevailing chronic disease management from individual stakeholder perspectives. The data mining sub module aims to extract patterns from

Table 1: Cluster Summary.

	E1	E2	E3	E4
gender	0.522	0.236	0.518	0.685
single	6.94E-019	0.0467	0.112	0
defacto	0.000429	0.0208	0.00577	7.38E-020
married	0.0259	0.384	0.324	7.32E-013
widowed	5.9E-018	5.52E-018	0.00121	0.977
divorced	0.974	0.000275	1.56E-005	0
separated	2.17E-019	0.0152	1.02E-019	7.38E-020
drinking status	0.998	0.816	0.738	0.29
smoking status	0.425	0.403	0.287	0.129
age	63.6	61	68.2	79.6
average HbA1c	**7.41**	**8.17**	**6.6**	**6.78**
minimum HbA1c	6.25	7.15	6.14	6.17
maximum HbA1c	10	9.55	7.19	7.48
standard deviation	1.58	1.04	0.414	0.445
% increase	**59.6**	**42.5**	**39.9**	**33.2**
% decrease	**23.9**	**46.5**	**39.2**	**31.7**
% stable	**15.7**	**10.8**	**20.6**	**33.5**
cluster quantization error	0.000349	1.3E-005	0.000116	2.39E-005

Table 2: Cluster Analysis.

Cluster	Patients's demographics and lifestyle	Average HbA1c	HbA1c variation
E1	divorced, elderly (average age of 64) males and females who are heavy drinkers and ex-smokers	> 7%	high tendency to increase
E2	males, considerable number of married and mostly with unknown marital status who are heavy drinkers and ex-smokers with an average age of 61	> 7%	Fluctuates heavily above normal
E3	elderly (average age of 68) males and females, considerable number of married and mostly with unknown marital status who are moderate drinkers and non smokers	< 7%	Stable
E4	very elderly (average age of 80) widowed females who neither drink nor smoke	< 7%	Stable

the cdmNet data which may potentially provide solutions to many open questions underlying chronic disease management.

The paper contains only the preliminary results obtained using the BI module. Our future work include enhancing the data mining module: (1) to include other data mining functionalities such as classification, prediction, association mining, and text mining; (2) to identify measures that determine healthcare outcomes in chronic disease management; and (3) to identify correlation among all variables found in the cdmNet data and the potential health outcomes.

Acknowledgements

References

Postgresql, 1996. URL http://www.postgresql.org/. Accessed: 01. March 2010.

Chronic disease management, 2007. URL http://www.adgp.com.au/site/index.cfm?display=333. Accessed: 12. November 2007.

Pentaho community wiki home, 2008. URL http://wiki.pentaho.com/display/COM/Community+Wiki+Home. Accessed: 01. March 2010.

Chronic disease management (cdm) medicare items, 2009. URL http://www.megpn.com.au/Docs/MBS%20items/Overview%2520Fact%2520Sheet.pdf.

Viscovery, 2010. URL http://www.viscovery.net/publications/books. Accessed: 01. May 2010.

G. Anderson and K. Wilson. Chronic disease in california: Facts and figures, 2006.

H. R. Z. T. Anderson, G. and N. Johnson. Chronic conditions: Making the case for ongoing care, 2004.

M. Berry and G. Linoff. *Data mining techniques: for marketing, sales, and customer relationship management.* Wiley New York, 2004.

U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy. *Advances in knowledge discovery and data mining.* MIT press, 1996.

M. Georgeff and J. Hilton. Why is telemedicine not more widely used? *Feature*, pages 32 – 41, 2010. URL www.pulseitmagazine.com.au.

M. Georgeff, J. Hilton, L. P. AM, C. Chapman, L. Segal, M. Paterson, P. Schattner, K. Jones, A. Adaji, D. Mortimer, T. D. AM, B. Costa, K. Fitzgerald, B. Rasmussen, and K. Sweeny. Cdm-net:a broadband health network for transforming chronic disease management. Final report, Precedence Health Care, 2010.

R. Kimball, M. Ross, and R. Merz. *The data warehouse toolkit: the complete guide to dimensional modeling.* Wiley, 2002. ISBN 0471200247.

T. Kohonen. Analysis of a simple self-organizing process. *Biological cybernetics*, 44(2):135–140, 1982a.

T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological cybernetics*, 43(1):59–69, 1982b.

B. Larson. *Delivering Business Intelligence with Microsoft SQL Server 2008.* McGraw-Hill Osborne Media, 2008.

S. Sibte and R. Abidi. Healthcare Knowledge Management: The Art of the Possible. 4924:1–20, 2008.

K. Smith. Introduction to Neural Networks and Data Mining: For Business Applications. *Journal-Operational Research Society*, 51(6):771–772, 2000.

E. Wagner, B. Austin, C. Davis, M. Hindmarsh, J. Schaefer, and A. Bonomi. Improving chronic illness care: translating evidence into action. *Health Affairs*, 20(6):64, 2001a.

E. Wagner, N. Sandhu, K. Newton, D. McCulloch, S. Ramsey, and L. Grothaus. Effect of improved glycemic control on health care costs and utilization. *Jama*, 285(2):182, 2001b.

Defining the Paperless Workplace with the Paper Metaphor - Not a Contradiction in Terms

Gerald Weber

Department of Computer Science
The University of Auckland
38 Princes Street, Auckland, New Zealand
Email: gerald@cs.auckland.ac.nz

Abstract

The transition from a paper-based work environment to a largely paperless environment is still in full swing, in healthcare as well as in other domains. Analysts predict a further decade of efforts is necessary at least. In reality, paperless IT-based workflows offer both advantages and disadvantages over paper-based solutions. This is in contrast to the naïve expectation that a paperless solution should be a strict improvement over paper-based processes. We identify a set of generic requirements that address common drawbacks of IT solutions, and we propose a system model that helps to create IT systems which preserve the advantages of paper-based processing. The main tenet is that the paperless solution should be based on a naturalistic paper metaphor. Our system model supports auditability of IT systems by direct reference to the paper metaphor and ensures that information is faithfully presented to the practitioner. The system model is intended for mission critical applications such as health record management.

1 Introduction

The transition from paper-based systems to largely paperless systems is a process that might still continue for many years. More conservative predictions for the healthcare sector (Ford et al. 2006) have come up with dates around 2024 for a thorough rollout of paperless solutions. The healthcare sector is well positioned in comparison to many other areas, not least because of a strong interest on the part of policy makers. New substantial stimulus packages have been created to achieve that goal, but at the same time they bear witness that we are not there yet.

As with other areas in IT, a plethora of reasons can cause the implementation of a paperless workflow to be riddled with problems. As a consequence, in some cases this means that the workflow becomes worse than with the paper predecessor. Immature or faulty underlying software packages, problems in process management or administration are possible reasons, although the impact of such problems is disputed (Glass 2006). However these are beyond the scope of our work. Our focus is on aspects of the system *architecture* that can either help or hinder the system to be better than the old style paper office. We want to capture these aspects of architecture not primarily as technical blueprints but by naming important concepts that the architecture must inherently

support.

It should be noted that one aspect making the transition harder is the fact that the translation from paper to IT-based work paradigms is not uniquely defined. We propose that careful observation of the needs of important domains can lead to stricter requirements for paperless workplace solutions. The central tenet will be that the system should incorporate aspects of the pre-IT workflows with a naturalistic and well-defined paper metaphor. The physical paper is turned into a virtual successor, initially with as few changes as possible. Ideally, the virtual paper should maintain all advantages of the physical paper, while offering the possibility to get rid of all drawbacks of physical paper. However, it is not always straightforward to decide if a property of paper is an advantage or a drawback. These considerations also help us in identifying aspects where current IT-based systems fall behind paper-based systems.

The paper is structured as follows. In Section 2 we highlight maintainability and auditability risks in current multitier applications that will be addressed throughout the paper. We present alternative system models that contain a naturalistic paper metaphor in Section 3. In Section 4 we present a model in which the naturalistic paper metaphor can be combined with traditional form-based interfaces. In Section 5 we review how this approach relates to other domains of IT use.

2 Advantages and Challenges of Paperless Systems

Our work is related to earlier work on Hippocratic databases (Agrawal et al. 2002), a type of advanced database applicable to privacy-intensive applications including healthcare applications. These systems create the possibility of auditing which data has been accessed at which point in time and in principle by whom. On the one hand Hippocratic databases support enforcement of privacy policies; on the other hand they support audits of compliance. This means that they have to reconstruct faithfully what actually happened. The compliance audit is based on an audit trail that stores database accesses. This enables some detailed analysis using access statistics that are available to the database. If we want to interpret these database accesses, however, we have to make some assumptions on the information system that processes the data before the presentation to the user. In particular we have to know the user's identity. If a database is used as a backend in a classical multi-tier architecture, the database has no reliable access to the actual user identity. The database access happens through artificial system user accounts that do not correspond to a single natural person. More generally, as soon as an architecture has a classical application server tier, this tier has unrestricted ability to transform

the data. It is not directly possible to infer from the database accesses what the users have actually seen and entered. In contrast, we present here a framework based on a naturalistic paper metaphor. The audit happens much higher up in the tier architecture, as close to the user interface as possible. The system keeps an audit trail of exactly what has been presented to the user and what the user has entered in which context. Still this can only be achieved under certain assumptions of correctness of the implementation of the framework. If the framework records the input incorrectly, this assumption is broken. This again highlights that the transition to an electronic system remains to some extent precarious. A further assumption must be made that all information rendered by the GUI is actually presented visually to the user and not, for example, occluded by other windows. Occlusion-free user interface technologies can help here (Lutteroth 2006).

Paperless systems allow the storage of information in presentation-independent form. The classical current implementation of this feature is the storage of data from paper-like input forms in databases. This feature will be a central topic of our subsequent discussion, where we will shed a critical light on it. This presentation-independent storage requires what is known as the shredding of information. The input forms are taken apart and the user-provided information is stored in a presentation-independent form. The unity of the information that was presented on one input screen is not only broken occasionally, in certain prevailing database design methodologies the data is preferably taken apart. Such shredding is also well known in XML processing (Amer-Yahia et al. 2004). Database normalization also asks for the decomposition of tables, if they are not in normal form (Codd 1971). However, following best practices (Brodie & Schmidt 1982), the data is then presented to higher layers in application specific virtual views. In a plain multitier architecture, there is no inherent guarantee from the system architecture that a value entered under, for example, the label "Age" is also presented under the same label in the system output. However, in our system model, this will be guaranteed. Such guarantees are also helpful in scenarios where more customized user interfaces could be used, for example in tele-health applications that are directed towards the end user (Singh et al. 2010).

This shredding of data, common in systems with relational database backends, immediately reinforces the aforementioned problem of long term data storage, since the shredding can amount to an obfuscation of the context and provenance of the data (Buneman & Tan 2007). Moreover the shredding creates problems for usability and auditability.

3 System Models with a Naturalistic Paper Metaphor

In contrast to system models that are based on a universal middle tier, we want to discuss models that present a naturalistic paper metaphor. As indicated in the introduction, we begin with a system that turns the physical paper into a virtual succcessor with as few changes in properties as possible. from that we move on to system models that include more and more of the additional features of an electronic system.

3.1 The Single Copy Model

Figure 1 shows a conceptual diagram of a paperless system that is one of the most direct translations of a paper-based system and that we call the *singleCopy*
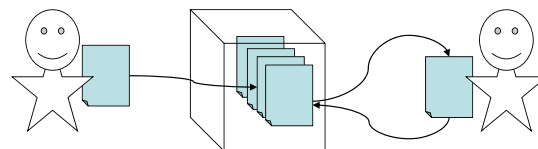

Figure 1: In the singleCopy model, each paper record is translated one-to-one into an electronic record. The user on the right hand side is taking out a record and putting it back into the old place within the repository.

model. Each paper record is translated one-to-one into an electronic record. The records can be deposited in the central repository and taken out on demand. In this system model we assume that the electronic records still inherit a fundamental property of paper-based records, namely the fact that copying requires an explicit operation. In this model, a record can only be taken out by one practitioner at a time. The immediate improvement that we obtain over paper-based records is the instantaneous communication with the repository, and possibly quick search functions. The fact that only one person at a time can have the record is on the first glance a disadvantage for, e.g., reading, but it can have several natural positive consequences. First of all this paradigm prevents write conflicts. People with write access will work strictly sequentially on the record. In addition, the paradigm can also naturally limit any prolific and inflationary access to personal data. If we assume that the record is kept with each practitioner only while actually needed, then this minimizes situations where concurrent access is needed. This kind of system functionality can be useful in several domains. In health care it might support the logic that the record travels with the client. In human resources it supports the concept that personal records are only accessed if needed. In general, the singleCopy model has the advantage that it assigns a clear responsibility at each point in time.

The singleCopy model is a concurrency model, in that several collaborators could have interleaving access to the document, each one returning it as soon as possible. Such a concurrency model is often called a check in/check out model. It can be seen as a pessimistic model in that it prevents conflicts rather than trusting that conflicts would be rare. The singleCopy model is distinct from typical check in/check out models, in that the latter treat the process as a locking of a central resource, and therefore do not directly use the paper metaphor, i.e. a document that literally travels between workers.

3.2 Concurrent Read Access

Technically, the singleCopy model can of course be extended easily to include a concurrent read, on top of the exclusive write, since concurrent reads in themselves do not create write conflicts. This bends the naturalistic paper metaphor, and can best be explained with a copy on demand process that we will elaborate below. At this stage, the idea is that concurrent readers would get an explicitly immutable copy of the document. It is noteworthy that this extension immediately raises a well known concurrency issue, namely the question of which version the concurrent readers should see. For many applications it might be best if the concurrent readers always see the most recent version, in the interest of early information dissemination. However in other applications it might be appropriate that only finalized and authorized versions are seen by concurrent readers.

a) Input:

PatientData

Name	John Doe
Age	41
NextOfKin	Jane Doe

submit

b) Output, change in affordances only:

PatientData

Name	John Doe
Age	41
NextOfKin	Jane Doe

submit

c) Output, change in representation:

PatientData

Patient   : John Doe (age: 41)
Next of Kin : Jane Doe

Figure 2: An input form and two different representations of the saved, immutable content.

We captured this requirement in an earlier discussion as the *Signed User Memo Viewpoint*, which requires that only signed and therefore final memos are distributed (Weber 2008). This approach then might require the possibility that one user can finalize and sign intermediate units of work without releasing control of the document.

We summarize the singleCopy model by comparing it against the goals of preserving the advantages of paper and reducing the disadvantages. The single-Copy model keeps the inherent advantage of physical paper with regard to conflict free writing. This advantage will become clearer when the alternative is discussed in the following copyOnDemand model. We have pointed out that in principle it is possible to introduce concurrent read access. One disadvantage of the physical paper is the possibility of loss through misplacement or destruction. This can of course be mitigated with an electronic system. An archiving component that keeps duplicates can be added.

3.3 Unity of Input and Output Representation

The strict application of the naturalistic paper metaphor requires that there is no difference between the form as it is presented while it is being filled out and the way it is represented when the entered data is recalled at a later stage. In this view, filling out a form should be directly equivalent to writing on a paper form. Hence, if this form needs to be accessed it should be represented in the same state as it was submitted. Most current form-based systems deviate rather strongly from this view, and this is typically directly coincidental with the process of shredding that we mentioned earlier.

However, we can also identify some motivation for changing the presentation. One plausible and justified reason to change the presentation relates to what is called a change of affordances. Before submission, the data is editable, and after submission it might be immutable. This is a change in affordance and justifies a change in representation. A faithful representation of entered data would use the same form layout, still presenting the data in the form fields, but having them represented non-editable. In Figure 2, the editable form is shown in a) and the immutable form is shown in b). The only change is a graying out of the form fields and submit button to indicate the change of affordance. The particular visual representation as grayed text is not our main concern and the illustration might show that graying is not an optimally readable solution, but the important part is that the change of affordance has just the character of an annotation.

In an electronic form, the text field is more than a reserved preferred space for information input; rather it is the only space where information can be entered. This is in contrast to paper forms, where technically every area of the form can be written to. It is appropriate to apply the notion of interface modes that refers to distinct changes in the reactivity of the interface. The notion of interface modes refers to temporal distinctions of interface behavior, such as when num lock is activated on the keyboard. But in an interface with a pointing device, cursor and focus, this temporal distinction is coupled with spatiality, namely with the position of the cursor. In this sense physical paper forms are modeless, i.e. they depend on the discipline of the writer to fill out the form neatly. In electronic forms the divide between editable fields and non-editable form area is much more fundamental; often these two areas are realized with strongly diverging technologies. In many form technologies, such as classical HTML forms as well as the more recent XForms, the labels of the form fields are indeed not a logical part of the form, but just peripheral text that is accidentally placed next to a form field.

This technological divide may well be in itself a contributing factor to the frequent change of representation between editable forms and the later representation of the entered data. Data will be entered in electronic forms, but presented in tables that no longer have the visual appearance of text boxes any longer. A possible such representation is shown in Figure 2 c). Just from the standpoint of a naturalistic paper metaphor, Version b) should be preferred over Version c). The change of affordance itself can be sometimes problematic in that it prevents practitioners to do important changes immediately. In principle, if a version control system is available, then one should present the form always in an editable mode, since earlier versions can be recalled. This discussion will be taken up again in Section 5 in particular with respect to wikis.

A natural consequence of this consideration is to base a paperless system on an underlying system that is closely based on the system model proposed in Figures 1 and 4, where the documents, including forms, on one level do not change appearance between different retrievals. With regard to our goal of preserving the advantages of paper and adding all possible support that an IT system can offer, we achieve the following. The representation of the form is still faithful to the way it was entered. However the system can be configured to prevent further editing of data, if this is wanted.

3.4 Historic Shift of Fundamental Form Semantics

It should be noted that the shift between the different form paradigms in paper forms and electronic forms also coincides with a shift in the apparent form semantics. One historic origin of paper forms, and a presumed root of the word "form", can be found in early legal documents, where standardized formulaic clauses were used with fill-in gaps for, e.g., the name of the person they apply to. Such a form might be a standardised contract. An example is a promissory note, an earlier and more perilous kind of an IOU. The interesting part here is that the semantics of the form must be completely expressed on the completed form. This means that after filling in the gaps, the completed form turns into a clause in natural language, although perhaps using some legal or other jargon. The effective power of the form, its semantics, is completely explainable from the resulting natural language clause without any recourse to its genesis as a fill-in-the-blanks text.
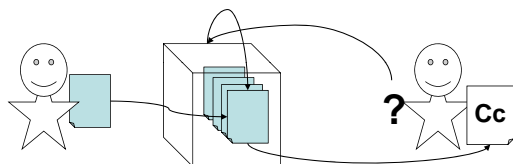
Figure 3: In the copyOnDemand model, if data is requested as indicated by the question mark, a new copy is created.

Figure 4: The copyOnDemand model allows parallel access to the same data, but can lead to conflicting updates.

In contrast, even before the advent of IT, many modern forms are designed, intended and presented as mere editable key/value pairs without any pretence to be self explanatory in their effect. The enterprise system behind the scene is the dominating element. The human readable elements of the form are merely making the required fields digestible to the untrained user and are semantically individual requests to divulge information: "Enter your name, enter your date of birth." This again coincides with the process of shredding the form data, and the fact that in a modern multitier information system only the filled-in fields are stored. Hence in a multi-tier application, even if the form fields were embedded in a fill-in-the-blanks text, this text would not have been recorded, and there would have been no tracking of the context in which the data was entered. From a human-computer interaction point of view this evolution is certainly not unproblematic.

It is worth noting that in certain critical applications such as bank cheques, which are incidentally closely related to promissory notes, the old style is still used and the filled in cheque contains a natural language clause. The aforementioned *Signed User Memo Viewpoint* is much closer to the earlier view of a form as a clause in natural language (Weber 2008). Under this viewpoint the form submission is separated into two distinct phases: first the form is prepared, this phase concluding with a pre-submit operation, and then the system presents the resulting memo that the user is about to authorise. This memo is not editable and it is made clear that the submission process is equivalent to the signature process. As a consequence the memo must be self-explanatory.

The modern key/value pair view is justified in circumstances where the form semantics is essentially data capture. Many healthcare applications fall into this category, for example if a vital parameter is regularly checked. However, the aforementioned problem remains: a multi-tier architecture offers no inherent guarantee that the data is presented under the same heading as it was entered. The shredding of information has another motivation in the recently increased interest in data integration. There is an interest to integrate data from different forms that might use slightly different field labels, into a single database. This requires a matching of equivalent fields. However, we recommend using a table style proposed in an earlier work (Lutteroth & Weber 2006) that keeps the original label with each data entry as seen in the following example:

last name:	first name:	phone:
Johnson	Kelly	0987654
family name:	given name:	phone:
Gray	Pat	0123456
surame:	first name:	ph:
Smith	Bo	3456789

3.5 Copy on Demand

We have gained the first system model, the single-Copy model, by a direct translation of paper proper-

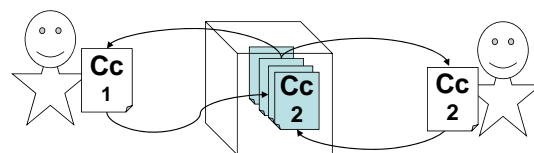ties to the IT world. We will now move away from that model in order to capture more of the possibilities of electronic media. However it should be noted upfront that the result of our considerations will be a model that is much closer to this initial model than to some of the more flexible general-purpose IT systems.

Many superficially exciting features of IT systems come with a price for serious, critical business processes. The first such property that turns out to be a mixed blessing for reasons already outlined is that paperless systems allow the parallel access to the same data. In Figure 3 we present a modified system model that we call the *copyOnDemand* model. Here records are obtained by copying information. This system model is a natural consequence of digital information processing, where it is simple to copy data, but it would require a deliberate effort to implement a restricted model such as the singleCopy model. Accordingly, the copyOnDemand model can have the problem of unrestricted proliferation of copies of the data with unclear semantics. In particular, version conflicts can arise if several people access the data at the same time, as shown in Figure 4. Sometimes such conflicts can be solved with automatic merge. Sometimes they require manual merge, however. This is hard enough for experienced software engineers; it would be naïve to assume that non-IT professionals can solve such problems, not because of a lack of understanding but because the necessary merge tools have a complex interface that an occasional user is not familiar with.

4 The Form-Oriented System Model

We will now introduce and apply the form-oriented system model (Draheim & Weber 2004), which is a simple yet powerful model for IT systems in the paperless office. This model will allow us to naturally address a number of the aforementioned issues and will ensure that at least the common drawbacks of paperless systems are mitigated. The model is easy to apply to all platforms and we plan to further explore its application to systems based on standardization efforts such as openEHR.

In the form-oriented system model, all possible input and output screens are defined by explicit screen prototypes that show the structure of information presented, as shown in Figure 5. The form-oriented methodology is designed for so-called submit-response style systems, a system class which includes classical Web interfaces, but also most proprietary enterprise application frameworks from individual vendors. The form-oriented system model is therefore suitable for a whole range of systems and can be used to improve the architecture of such systems. The model can, for example, be applied to systems based on Java EE. The model is, however, equally applicable to other technologies such as commercial of-the-shelf enterprise software and can be used to provide better system documentation, for example in the context of a comprehensive quality management.
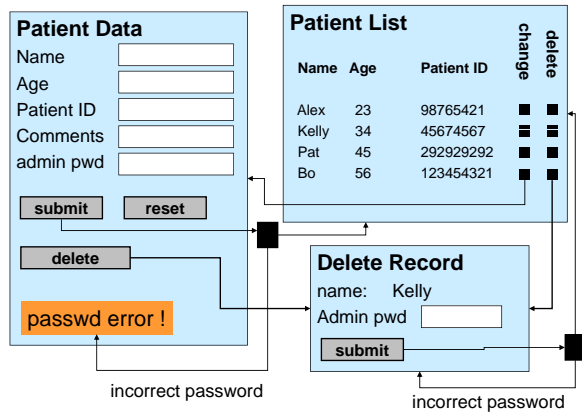
In the past, a focus of form-oriented system models

Figure 5: A form-oriented screen diagram, showing prototypes of the actual system screens and state transitions between them.

was on defining the flow of dialogues, and the necessary state machine defining the dialogue. One such state machine is represented in Figure 5 by the rectangular elbow arrows connecting the screens. The diagram in Figure 5 is called a screen diagram. It is a largely self-explanatory type of diagram that can be refined in this methodologyto more formal models such as bipartite formcharts. In our discussion here we will leave the flow of dialogue screens aside and focus on the individual dialogue screens and their mutual consistency.

The form-oriented model provides a synthesis of the singleCopy model and the copyOnDemand model in that it adds structure to the version space of Figure 3 yet matches well the natural structure of IT systems. The form-oriented model takes into account that in critical applications, data updates have to be done consciously and carefully. For that purpose the system interaction is modeled with traditional input forms as they are encountered on paper as well as, for example, in web forms. The filling out of the form is viewed as a preliminary activity that has no durable semantics in itself and only serves the purpose of preparing the actual submission of the data. The submission of the data happens through an atomic operation, typically hitting a button, which should only happen after carefully reviewing the data.

This model delivers a two-stage interaction paradigm, the preliminary preparation of the data and the atomic, heavyweight submission of the data. This matches a traditional paper-based work pattern, where people prepare and submit forms. In an IT system context it is, however, natural that after the submission the system continues in a specific state. The dialogue with the system is a continuous alternation of screens offered by the system and user interactions with the screen. Therefore the submission process in our model is called a page change, and after submission the user finds the next set of choices. For simplicity's sake, browsing actions of the user are modeled in the same paradigm as memo preparation. Hence in the form-based model, filling in a search screen is modeled as the same type of form as the submission of a professional assessment.

All these forms will be handled and archived in the same way. The form-oriented system model enables a system architecture, where all the important information is archived in the shape in which it was exchanged between user and system. This gives maximum flexibility for dealing with the system in a uniform way, yet still offers a very easy way to ignore unimportant data, such as all search forms, should they be irrelevant for the task at hand. The storage
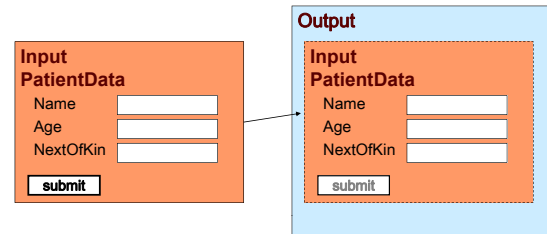

Figure 6: The form oriented system model allows the simple literal reuse of input data in subsequent output.

of all input forms in an audit trail still allows the on-demand blackening of sensitive data, thus enabling limited disclosure (LeFevre et al. 2004).

4.1 Usability Problems of Shredded Data and Solutions

In our research we are interested in crosscutting aspects of usability in IT systems (Weber 2008). One focus is on usability problems that create what is known as a gulf of evaluation (Norman & Draper 1986). A typical functional requirement of an e-health system would be, for example, that lab data for a client is captured and stored. This however does not guarantee availability of the data. A sensible requirement would be that all lab data that is created at a single point in time can also be viewed as a unit in the future. We call this a faithfully reporting system. This would be a direct consequence of the paper metaphor; the lab report is preferably always presented in the same, complete version. This avoids a possible gulf of evaluation, namely that the practitioner sees one lab value, but cannot necessarily see other lab data or even ascertain which tests were done at the same time. This example is a more complicated requirement than the aforementioned requirement that one lab value should be always presented with the same label. We will now discuss how far this requirement is addressed in a plain multi-tier architecture, and whether it is addressed in the form-oriented architecture.

Plain multi-tier systems store data in a database backend that can be used by a variety of applications. These applications in turn present themselves to the user through input and output screens which can be designed completely independently (Brodie & Schmidt 1982). With the flexibility comes a risk. In the fundamental system model of multi-tier applications there is no guarantee that data that was entered at the same time is also presented in the output at the same time. This is a direct consequence of shredding, as defined above. Hence in the plain multi-tier architecture a system design can easily be created where this gulf of evaluation arises. Since the presentation of the data is completely configurable, selected data can be presented, which might seem sensible to the developer, but leads to problems for the practitioner who uses the system. Avoiding the aforementioned gulf of evaluation requires an effort, namely active and adequate checking by the developers that the right information is provided at every place. Hence in plain multi-tier architectures it is harder to create a faithfully reporting model which will follow the terminology used before.

In the form-oriented system model, as opposed to the multi-tier model, faithful representation of the data as it was entered is always available. This is shown in Figure 6, where on the left side an input screen for client data is shown. The figure shows an example of a form-oriented model of the screens in a

very minimal system. This model, however, already captures one feature of a realistic system, namely corresponding input and output. In this model the output screen is shown on the right-hand side. As we see, we can define this output screen in the form oriented methodology by reusing the input screen one-to-one in the output. If, as shown in this example, only a single input screen is quoted, we would not need this representation, but could use the representation used in Figure 2 b), but this example serves as a demonstration for the more complex reuse examples that will be discussed in due course. The submit button of the input screen is disabled, but the content is represented faithfully. This model has a rigorous formal underpinning in the semantic model of form-oriented analysis (Draheim & Weber 2004) but it is also intuitively understandable. The arrows denoting reuse are always shown as straight arrows that are neither horizontal nor vertical, to distinguish them from the rectangular elbow arrows indicating state transitions.

This system model addresses the problem of shredding. It is now simpler to create a faithfully reporting model than a different, shredded model, using the terminology from before. Further questions arise now with respect to the long-term consistency of the model over time, and we will address them in the following by proposing suitable modeling solutions. In that way the reuse concept that we have applied in Figure 6 can be extended to a system design methodology that allows flexible creation of rich and interesting user interfaces, yet mitigates the problem of shredding. Shredding will not be completely prevented, since this would mean that the modeling method would cease to be universal. We deem it sufficient that a shredded model would be more complex than a non-shredded model in a metric that counts modeling elements.

4.2 Reuse of Screens in rich User interfaces

The reuse of input screens in output screens is the most basic and most fundamental example of a system that relies on the faithful representation of data. The form-oriented models are suitable as models to be developed in a standard graphical editor. Assuming that Figure 6 was created in a standard graphical editor, the question arises as to what keeps the model consistent if the definitions of the input forms are changed. We therefore use a different graphical representation in form-oriented models that makes the reuse explicit by avoiding the duplication of model elements. This is shown in Figure 7. The part of the input screen that is reused is called a snippet. The snippet is not repeated at the place where it is applied, but a placeholder is entered that is connected with the original snippet. The figure also shows that the definition of snippets can be nested. The semantics of this notation is that changes to the original snippet apply immediately and consistently to all other applications of the snippet. In this way we have achieved the long-term guarantee of a faithfully reporting system.

From the perspective of a naturalistic paper metaphor the snippet character should be visibly retained so that the end user can see that the snippet is part of a possibly wider input form. A more comprehensive solution could be found by applying the technology of transclusion in a document-oriented approach that we explored earlier (Lee et al. 2010). The notation in Figure 7 uses a special notation for selections. In order to change or delete a record the user of the system as shown in Figure 5 could click a button next to the person. In principle this can be supported also with our new approach. We choose here to use a slightly different but interesting approach. Every patient record has a small handle indicated with a
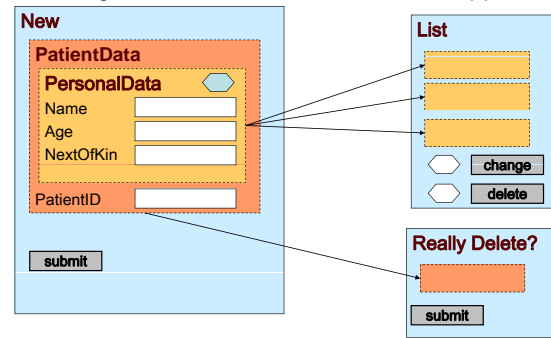

Figure 7: Non-redundant reuse ensures that there is a single place of maintenance for snippets.
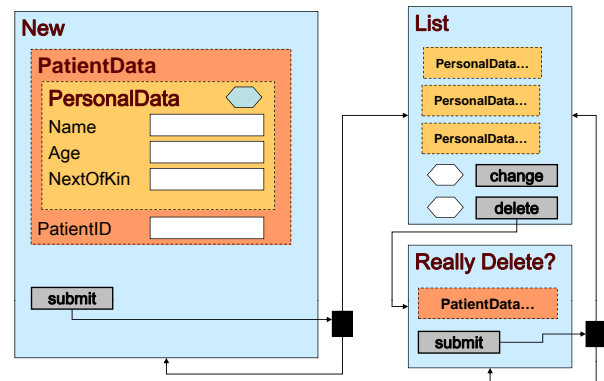

Figure 8: The semantics of the snippet reuse is symbolic, and a symbolic notation can scale up to large, multipage system models. The sympolic notation is used here within a screen diagram.

hexagonal shape. This small handle is a symbolic representation of the id of this record. In the system represented in Figure 7 the user can drag and drop this handle into a form field of the same shape, akin to a simple jigsaw puzzle. The two forms for change and delete have such a matching form field each. The small handles that enable drag-and-drop we call *opaque references* in form-oriented analysis (Draheim & Weber 2004, Draheim et al. 2005).

It is important to note that the reuse is not a purely graphical notation. In large systems, graphical models are often not scalable. The semantics underlying this reuse model is symbolic and this is shown in Figure 8. The snippets are addressed with symbolic names, and in the spirit of semantic faithfulness, by default this is the screen title of a snippet. Hence the Snippet "PersonalData" is invoked by using that name in a placeholder element, followed by the elision operator "...".

The symbolic notation makes it particularly easy to combine the model with the form oriented model of page change. This is also shown in Figure 8. The screen diagram also highlights the semantics of opaque references: on the right side, we see that after submitting the delete form on the list page we are taken to the delete page.

The reuse of snippets already used in other places is an intuitive first step to more maintainability. In large projects, the reuse model as shown in Figure 8 would, however, still exhibit an asymmetry. The maintenance of the snippet is tied to the first page where it is used, or at least the current page where it is placed at a given point in time. In a large-scale system development it is worth-while to use the snippet approach to perform a separation of concerns. The
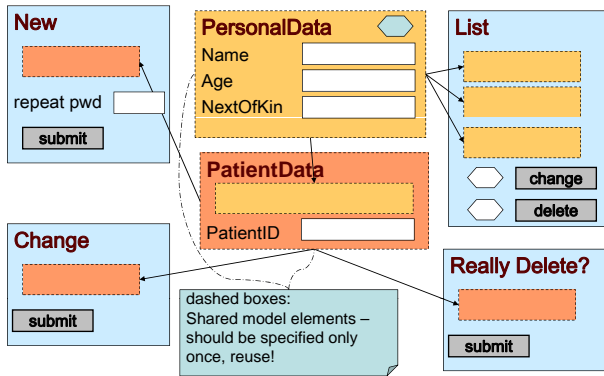
Figure 9: A reuse model that distinguishes between the shared model elements and the actual input and output.

snippets that are reused are maintained in a common shared model, and are kept separate from the pages where they appear. All placements on actual input and output screens use placeholders and hence are treated equally. This is shown in Figure 9.

4.3 Risks in Bulk Storage of Sensitive Data

Paperless systems allow easier duplication through archiving for disaster recovery. As a consequence, however, paperless systems harbour the danger that complete copies of sensitive data can be lost or misused. Sensitive health records for a whole population fit onto a memory stick of the size available at corner newsagents. Several instances of accidental loss of data on thousands of childcare claimants, social insurance holders, law enforcement officials, and other cases have been documented (BBC 2008) and the possibility of deliberate misuse of personal data for financial gain looms large.

Although paperless systems allow disaster-safe archiving of records, the long term storage is problematic. All currently available storage media are exposed to several independent long term risks: the accessibility of the data formats is not guaranteed, since the access to the information is highly indirect and depends in every case on a whole stack of formats. Furthermore currently available physical digital high-density storage media have a much shorter life expectancy than paper. Crucially for personal healthcare records, the life expectancy of digital media is typically shorter than the average human lifespan, while paper records have an established life expectancy longer than the average human lifespan. Experts in the field of digital data curation (Buneman et al. 2008) feel compelled to recommend explicit occasional paper copies of all sensitive data, e.g. printouts of XML data, a somewhat ironic state of affairs in the transition to a paperless process. Microfilm remains accepted as a higher density medium with similar life expectancy to paper.

5 Relation to other Domains

There is increasing interest in Electronic Laboratory Notebooks (ELN) which are likely to replace the traditional notebook, for example, in a research setting (Kihln 2005). These ELN are intended to help with IP disputes many years after the entries have been made. It is a requirement that ELN records are guaranteed authentic and tamper-proof, but also presented in a human-readable form (Drake 2007). This matches closely our considerations for faithful representation of the original entries, even if the main mo-

tivations might be slightly different. For a lab-book, the singleCopy system model is a natural first model. Two competing challenges are that, on the one hand, instant notetaking should be supported and therefore a modeless interface is intended. On the other hand it might be desirable that data series can be stored, and such data series are best represented in a tabular form.

A system model that comes very close to our goals can be found in Wikis. In particular the set of functionality provided by the MediaWiki platform is widely used now. Such a wiki supports on the one hand a concept of pages with identities that is close to the singleCopy model, in that there can be only one current version of the page. On the other hand it supports a copyOnDemand style approach to editing, since all the previous versions of a page are kept discoverable in the page history. There is a certain difference in that the singleCopy model has no checkout, thus Wikis are clearly an optimistic model. This model has some obvious advantages as well.

Another area which is concerned with reducing the divide between paper based workflows and electronic workflows is that of pen-based input. Pen-based systems approach the problem from the side of input devices, in enabling a style of interaction that is very close to working with paper, but is also enhanced by the new possibilities of electronic processing, including handwriting and drawing recognition (Plimmer & Apperley 2007).

In general, the problems addressed here are common to form-based systems and are encountered for example in more commercial applications. A classical case study is Electronic Data Interchange (EDI) (Emmelhainz 1992) that is widely used in the commercial world even before the advent of web-based e-commerce (Alonso et al. 2004).

This initiative has been valuable for retailers and suppliers since it allows them the electronic exchange of daily orders and business communication (Dowdeswell & Lutteroth 2005). The natural inflexibility of the automated systems contributed to a strong standardization of business communication. Only a few rigorously defined business message types are allowed (Kimberley 1991) in those systems, which contributed to interoperability between internationally distributed participants. In the healthcare area, openEHR is an example of a similar evolution. In the wake of the electronification of health records we observe a strong incentive to standardize health records on a global scale (Eichelberg et al. 2005). In a somewhat daring generalization from these two example observations we conjecture that there is a natural connection between the electronification of a business domain and the creation of stricter standard interchange formats.

6 Conclusion

The era of paper-dominated bureaucracy might well be in decline. The IT systems that replace paper are now a part of our everyday life. Once the initial excitement about the new tools has waned, users will expect rigorous quality attributes from such systems. IT systems are not automatically superior to paper-based solutions. Sometimes paper-based solutions come with natural and intuitive semantics. If push comes to shove, people are always able to perform at least basic functions with a stack of paper, such as browsing through all pages and looking out for headlines. In computer systems, even these basic functions have to be learned, and differ on every one of the competing systems at least.

In this work we have made a case that a naturalistic paper metaphor should be part of mission critical IT systems. We have focused on advanced parts of this approach, including dealing with snippets to make the various screens in a system consistent. We have discussed arguments why a system model that is close to the paper metaphor can foster better usability. A naturalistic paper metaphor can establish a well-defined abstraction layer; the technology that establishes the paper metaphor is separated from the application that uses the metaphor. The aim is to make the transition to a paperless system easy and intuitive for the practitioner, and hence give everyone working in important areas such as healthcare a bit more time to focus on what is most important, namely the professional service to the client.

References

Agrawal, R., Kiernan, J., Srikant, R. & Xu, Y. (2002), Hippocratic databases, *in* 'VLDB', Morgan Kaufmann, pp. 143–154.

Alonso, G., Casati, F., Kuno, H. & Machiraju, V. (2004), *Web Services: Concepts, Architecture and Applications*, Springer Verlag.

Amer-Yahia, S., Du, F. & Freire, J. (2004), A comprehensive solution to the xml-to-relational mapping problem, *in* 'WIDM '04: Proceedings of the 6th annual ACM international workshop on Web information and data management', ACM, New York, NY, USA, pp. 31–38.

BBC (2008), 'Extent of data losses is revealed', *BBC News - Politics, 21st August* .

Brodie, M. L. & Schmidt, J. W. (1982), 'Final report of the ansi/x3/sparc dbs-sg relational database task group', *SIGMOD Rec.* **12**(4), 1–62.

Buneman, P., Cheney, J., Tan, W.-C. & Vansummeren, S. (2008), Curated databases, *in* 'PODS '08: Proceedings of the twenty-seventh ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems', ACM, New York, NY, USA, pp. 1–12.

Buneman, P. & Tan, W.-C. (2007), Provenance in databases, *in* 'SIGMOD '07: Proceedings of the 2007 ACM SIGMOD international conference on Management of data', ACM, New York, NY, USA, pp. 1171–1173.

Codd, E. F. (1971), 'Further normalization of the data base relational model', *IBM Research Report, San Jose, California* **RJ909**.

Dowdeswell, B. & Lutteroth, C. (2005), A message exchange architecture for modern e-commerce., *in* D. Draheim & G. Weber, eds, 'TEAA', Vol. 3888 of *Lecture Notes in Computer Science*, Springer, pp. 56–70.

Draheim, D., Lutteroth, C. & Weber, G. (2005), Robust content creation with form-oriented user interfaces, *in* 'Proceedings of CHINZ 2005 - 6th International Conference of the ACM's Special Interest Group on Computer-Human Interaction', ACM Press.

Draheim, D. & Weber, G. (2004), *Form-Oriented Analysis - A New Methodology to Model Form-Based Applications*, Springer.

Drake, D. J. (2007), 'Eln implementation challenges', *Drug Discovery Today* **12**(15-16), 647 – 649.

Eichelberg, M., Aden, T. & Riesmeier, J. (2005), 'A survey and analysis of electronic healthcare record standards', *ACM Computing Surveys* .

Emmelhainz, M. A. (1992), *EDI: Total Management Guide*, John Wiley & Sons, Inc., New York, NY, USA.

Ford, E. W., Menachemi, N. & Phillips, M. T. (2006), 'Predicting the Adoption of Electronic Health Records by Physicians: When Will Health Care be Paperless?', *Journal of the American Medical Informatics Association* **13**(1), 106–112.

Glass, R. L. (2006), 'The standish report: does it really describe a software crisis?', *Commun. ACM* **49**(8), 15–16.

Kihln, M. (2005), 'Electronic lab notebooks - do they work in reality?', *Drug Discovery Today* **10**(18), 1205 – 1207.

Kimberley, P. (1991), *Electronic Data Interchange*, McGraw Hill.

Lee, L.-C., Lutteroth, C. & Weber, G. (2010), Improving end-user gui customization with transclusion, *in* B. Mans & M. Reynolds, eds, '33rd Australasian Computer Science Conference (ACSC 2010)', Vol. 102 of *CRPIT*, ACS, Brisbane, Australia, pp. 163–172.

LeFevre, K., Agrawal, R., Ercegovac, V., Ramakrishnan, R., Xu, Y. & DeWitt, D. J. (2004), Limiting disclosure in hippocratic databases, *in* M. A. Nascimento, M. T. Özsu, D. Kossmann, R. J. Miller, J. A. Blakeley & K. B. Schiefer, eds, 'VLDB', Morgan Kaufmann, pp. 108–119.

Lutteroth, C. (2006), AP1: A platform for model-based software engineering, *in* D. Draheim & G. Weber, eds, 'TEAA', Vol. 4473 of *Lecture Notes in Computer Science*, Springer, pp. 270–284.

Lutteroth, C. & Weber, G. (2006), User interface layout with ordinal and linear constraints, *in* 'AUIC '06: Proceedings of the 7th Australasian User Interface Conference', Australian Computer Society, Darlinghurst, Australia, Australia, pp. 53–60.

Norman, D. A. & Draper, S. W. (1986), *User centered system design: new perspectives on human-computer interaction*, Lawrence Erlbaum Associates, Hillsdale.

Plimmer, B. & Apperley, M. (2007), Making paperless work, *in* 'Proceedings of the 7th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction: design centered HCI', CHINZ '07, ACM, New York, NY, USA, pp. 1–8.

Singh, J., Wünsche, B. & Lutteroth, C. (2010), Framework for Healthcare4Life: a ubiquitous patient-centric telehealth system, *in* 'Proceedings of the 11th International Conference of the NZ Chapter of the ACM Special Interest Group on Human-Computer Interaction on ZZZ', ACM, pp. 41–48.

Weber, G. (2008), A platform-independent approach for auditing information systems, *in* 'HDKM '08: Proceedings of the second Australasian workshop on Health data and knowledge management', Australian Computer Society, Inc., Darlinghurst, Australia, Australia, pp. 65–73.

Author Index

Recent Volumes in the CRPIT Series

Listed below are some of the latest volumes published in the ACS Series *Conferences in Research and Practice in Information Technology*. The full text of most papers (in either PDF or Postscript format) is available at the series website `http://crpit.com`.

Volume 91 - Computer Science 2009
Edited by Bernard Mans Macquarie University. January, 2009. 978-1-920682-72-9.

Contains the proceedings of the Thirty-Second Australasian Computer Science Conference (ACSC2009), Wellington, New Zealand, January 2009.

Volume 92 - Database Technologies 2009
Edited by Xuemin Lin, University of New South Wales and Athman Bouguettaya, CSIRO. January, 2009. 978-1-920682-73-6.

Contains the proceedings of the Twentieth Australasian Database Conference (ADC2009), Wellington, New Zealand, January 2009.

Volume 93 - User Interfaces 2009
Edited by Paul Calder Flinders University and Gerald Weber University of Auckland. January, 2009. 978-1-920682-74-3.

Contains the proceedings of the Tenth Australasian User Interface Conference (AUIC2009), Wellington, New Zealand, January 2009.

Volume 94 - Theory of Computing 2009
Edited by Prabhu Manyem, University of Ballarat and Rod Downey, Victoria University of Wellington. January, 2009. 978-1-920682-75-0.

Contains the proceedings of the Fifteenth Computing: The Australasian Theory Symposium (CATS2009), Wellington, New Zealand, January 2009.

Volume 95 - Computing Education 2009
Edited by Margaret Hamilton, RMIT University and Tony Clear, Auckland University of Technology. January, 2009. 978-1-920682-76-7.

Contains the proceedings of the Eleventh Australasian Computing Education Conference (ACE2009), Wellington, New Zealand, January 2009.

Volume 96 - Conceptual Modelling 2009
Edited by Markus Kirchberg, Institute for Infocomm Research, A*STAR, Singapore and Sebastian Link, Victoria University of Wellington, New Zealand. January, 2009. 978-1-920682-77-4.

Contains the proceedings of the Fifth Asia-Pacific Conference on Conceptual Modelling (APCCM2008), Wollongong, NSW, Australia, January 2008.

Volume 97 - Health Data and Knowledge Management 2009
Edited by James R. Warren, University of Auckland. January, 2009. 978-1-920682-78-1.

Contains the proceedings of the Third Australasian Workshop on Health Data and Knowledge Management (HDKM 2009), Wellington, New Zealand, January 2009.

Volume 98 - Information Security 2009
Edited by Ljiljana Brankovic, University of Newcastle and Willy Susilo, University of Wollongong. January, 2009. 978-1-920682-79-8.

Contains the proceedings of the Australasian Information Security Conference (AISC 2009), Wellington, New Zealand, January 2009.

Volume 99 - Grid Computing and e-Research 2009
Edited by Paul Roe and Wayne Kelly, QUT. January, 2009. 978-1-920682-80-4.

Contains the proceedings of the Australasian Workshop on Grid Computing and e-Research (AusGrid 2009), Wellington, New Zealand, January 2009.

Volume 100 - Safety Critical Systems and Software 2007
Edited by Tony Cant, Defence Science and Technology Organisation, Australia. December, 2008. 978-1-920682-81-1.

Contains the proceedings of the 13th Australian Conference on Safety Critical Systems and Software, Canberra, Australia, December, 2008.

Volume 101 - Data Mining and Analytics 2009
Edited by Paul J. Kennedy, University of Technology, Sydney, Kok–Leong Ong, Deakin University and Peter Christen, The Australian National University. November, 2009. 978-1-920682-82-8.

Contains the proceedings of the 8th Australasian Data Mining Conference (AusDM 2009), Melbourne, Victoria, Australia, November, 2009.

Volume 102 - Computer Science 2010
Edited by Bernard Mans, Macquarie University, Australia and Mark Reynolds, University of Western Australia, Australia. January, 2010. 978-1-920682-83-5.

Contains the proceedings of the Thirty-Third Australasian Computer Science Conference (ACSC 2010), Brisbane, Queensland, Australia, January 2010.

Volume 103 - Computing Education 2010
Edited by Tony Clear, Auckland University of Technology, New Zealand and John Hamer, University of Auckland, New Zealand. January, 2010. 978-1-920682-84-2.

Contains the proceedings of the Twelfth Australasian Computing Education Conference (ACE 2010), Brisbane, Queensland, Australia, January 2010.

Volume 104 - Database Technologies 2010
Edited by Heng Tao Shen, University of Queensland, Australia and Athman Bouguettaya, CSIRO ICT Centre, Australia. January, 2010. 978-1-920682-85-9.

Contains the proceedings of the Twenty-First Australasian Database Conference (ADC 2010), Brisbane, Queensland, Australia, January 2010.

Volume 105 - Information Security 2010
Edited by Colin Boyd, Queensland University of Technology, Australia and Willy Susilo, University of Wollongong, Australia. January, 2010. 978-1-920682-86-6.

Contains the proceedings of the Eight Australasian Information Security Conference (AISC 2010), Brisbane, Queensland, Australia, January 2010.

Volume 106 - User Interfaces 2010
Edited by Christof Lutteroth, University of Auckland, New Zealand and Paul Calder Flinders University, Australia. January, 2010. 978-1-920682-87-3.

Contains the proceedings of the Eleventh Australasian User Interface Conference (AUIC2010), Brisbane, Queensland, Australia, January 2010.

Volume 107 - Parallel and Distributed Computing 2010 2010
Edited by Jinjun Chen, Swinburne University of Technology, Australia and Rajiv Ranjan, University of New South Wales, Australia. January, 2010. 978-1-920682-88-0.

Contains the proceedings of the Eigth Australasian Symposium on Parallel and Distributed Computing (AusPDC 2010), Brisbane, Queensland, Australia, January 2010.

Volume 108 - Health Informatics and Knowledge Management 2010
Edited by Anthony Maeder, University of Western Sydney, Australia and David Hansen, CSIRO Australian e-Health Research Centre, Australia. January, 2010. 978-1-920682-89-7.

Contains the proceedings of the Fourth Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2010), Brisbane, Queensland, Australia, January 2010.

Volume 109 – Theory of Computing 2010
Edited by Taso Viglas, University of Sydney, Australia and Alex Potanin, Victoria University of Wellington, New Zealand. January, 2010. 978-1-920682-90-3.

Contains the proceedings of the Sixteenth Computing: The Australasian Theory Symposium (CATS 2010), Brisbane, Queensland, Australia, January 2010.

Volume 110 – Conceptual Modelling 2010
Edited by Sebastian Link, Victoria University of Wellington, New Zealand and Aditya Ghose, University of Wollongong, Australia. January, 2010. 978-1-920682-92-7.

Contains the proceedings of the Seventh Asia-Pacific Conference on Conceptual Modelling (APCCM2010), Brisbane, Queensland, Australia, January 2010.

Volume 112 – Advances in Ontologies 2009
Edited by Edited by Thomas Meyer, Meraka Institute, South Africa and Kerry Taylor, CSIRO ICT Centre, Australia. December, 2009. 978-1-920682-91-0.

Contains the proceedings of the Australasian Ontology Workshop 2009 (AOW 2009), Melbourne, Australia, December, 2009.