International Journal of Computational Engineering Research||Vol, 03||Issue, 10||

Framework for Data Mining In Healthcare Information System in Developing Countries: A Case of Tanzania

[1,]Salim Amour Diwani, [2,]Anael Sam
School of Computational and Communications Science and Engineering, Nelson Mandela-African Institution of Science and Technology (NM-AIST), Arusha, Tanzania.

ABSTRACT :
Globally the healthcare sector is abundant with data and hence using data mining techniques in this area seems promising. Healthcare sector collects huge amounts of data on a daily basis. Transferring data into secure electronic system of medical health can save lives and reduce the cost of healthcare services as well as early discovery of contagious diseases with advanced collection of medical data. In this study we have proposed a best fit for data mining techniques in healthcare based on a case study. The proposed framework aims to provide self healthcare treatments where by several monitoring equipments using the cyberspace devices have been developed to help patients manage their medical conditions at home for example, diabetic patients can test their blood sugar level by using e-device, which ,with the click of a computer mouse, downloads the results to a healthcare practitioner, minimizes time to wait for medical treatments, and minimizes the delay time in providing medical treatments. Data mining is a new technology used in different types of sectors to improve the effectiveness and efficiency of business model as well as solving problems in business world.

KEYWORDS : *Healthcare, Data Mining, Knowledge Discovery, OLAP, NIMR, NGOs, DSS, KDD.*

I. INTRODUCTION

Medical data are highly complex and difficult to analyze where as financial data are well organized but pose limited clinical value. Clinical data are very poor from the point of view of automated analysis systems that collect high quality data which will become part of routine clinical care, but are unlikely to have a large patient impact in 5-10 years. In most cases medical data is highly complex and difficult to analyze while financial data is well organized but has limited clinical value. Since the gap is between data gathering and comprehension, this paper proposes the way to fill the gap in Tanzanian context. The proposed framework can be used to predict future medical conditions for deadly diseases occurring in Tanzania.

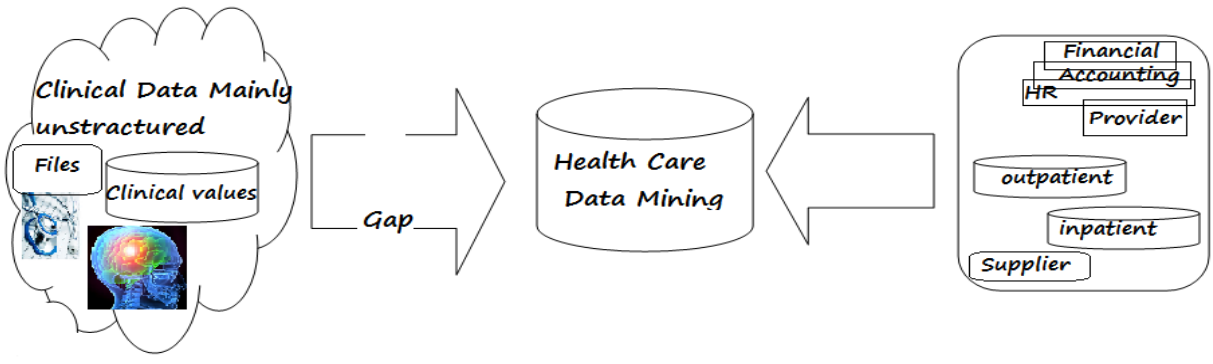

Figure 1: Block diagram capturing data Gap

Take for example how Netflix recommends movies and TV shows or how Amazon.com suggests products to buy. The framework makes predictions on what a patient has already experienced as well as the experience of other patients showing serious medical history. This provides physicians with insights on what might come next for a patient based on experiences of other patients. It also gives a prediction that is interpretable by patients.

The proposed framework can share information across patients who have similar health problems. This allows for better predictions when details of a patient's medical history are sparse. Data mining is an emerging technology used in different types of organizations to improve the efficiency and effectiveness of business processes. The application of data mining technologies would be of great benefit in assembling the required information, for example, in increasing operational efficiencies, fraud detection and enhance the overall decision making in organizations including public sectors **[1,2].** Data mining techniques analyze large data sets to discover new relationships between the stored data values. Healthcare is an information rich industry, warehousing large amount of medical data. The health-care industry finds it difficult to manage and properly utilize the huge medical data collected through different medical processes. Stored medical data collection is an asset for healthcare organizations if properly utilized. The healthcare industry can use data mining techniques to fully utilize the benefits of the stored medical datasets.

II. PROBLEM AND RELATED WORK

There is a lack of knowledge of the status of implementation of data mining technology within the healthcare system in Tanzania, the benefits of implementing such technologies and identification of best fit framework. Medical data mining is a key technique used to extract useful clinical knowledge from medical records. A number of scoring systems exist around the globe that use medical knowledge for various conditions but we don't have any in Tanzania. We have number of examples which uses data mining for various reasons:

- Arkansas data network evaluates re-admission and resources utilization, compares the data against current scientific literature and then determines the best treatments to lower spending [3].
- Group health co-operative sorts its patients by their demographic traits and medical conditions in order to discover which groups use the most resources. In this way, programs can be developed to help educate "problem" populations on how to better prevents or manage their conditions [3].
- The Acute Physiology and Chronic Health Evaluation (APACHE) series of models are developed to predict the individual patient's risk of hospital death in ICU, based on a number of physiological variables. The original APACHE model was developed in 1981 as an export-based scoring system. The later versions are based on logistic regression models. The models were trained on 17000 of cases in more than 40 hospitals [4].
- The Pneumonia Severity of Illness Index is another logistic regression model that predicts the risk of death within 30 days for adult patients with pneumonia. The model was developed by the Pneumonia Patient Outcome Research Team (PORT) in 1997 and was validated over 50000 patients in 275 hospitals in US and Canada. The developers claim that by using this model, up to 30% of pneumonia patients can be treated safely as outpatients, resulting in an annual savings of 1.2 billion dollars[4].
- investigation of the possible effects of multiple drug exposures at different stages of pregnancy on preterm birth, using Smart Rule, a data mining technique for generating associative rules [5].
- framework for video mining in vivo microscopy images to track leukocytes in order to predict inflammatory response which allows researchers to capture images of the cellular and molecular processes in a living organism [6].
- data mining based decision tools for evaluating treatment choices for uterine fibroids. The tool use data mining techniques to predict treatments choice for fibroids[7].

III. DATA MINING

Data mining uses a variety of techniques to find hidden patterns and relationships in large pools of data and infer from them that can predict future behaviors and guide in decision making [8]. Individuals and organizations are recognizing that addition value lie within the vast amount of data that they store. By applying data mining techniques, which are elements of statistics, artificial intelligence and machine learning; organizations are able to identify trends within the data they did not know existed. Data mining is a step in the knowledge discovery in databases (KDD) process and refers to algorithms that are applied to extract patterns from the data. The extracted information can then be used to form a prediction or classification model, identify trends and associations, refine an existing model, or provide a summary of the database being mined [9]. The output of a data mining exercise can take the form of patterns, trends or rules that are implicit in the data. Through data mining and the new knowledge it provides, individuals are able to leverage the data to create new opportunities or value for their organizations. Data mining is the activity of extracting data obtained from a variety of sources, usually held in a central data warehouse, for evaluation to assist in responding to questions posed, for example, by management. Data mining is a technical term that can be explained in terms of an individual's everyday life experiences;

we constantly extract data or information through our experiences and make decisions regarding our activities based on this information. In technological terms, the concept of data mining is known as the process of discovering new, valuable information from a large collection of raw data **[10,11,12]**and should enable better decision making throughout an organization **[13,14,15].** Because the architecture of a data mining model integrates various techniques and fields, it has meant different things to different people and it is not surprising that different ways of looking at the concept have taken place.

IV. Proposed Framework

A proposed framework for Tanzania healthcare can be developed and grouped into four categories: infrastructure, administrative, financial and clinical applications. In the proposed framework two web portals can be developed: one for the clinician and the other one for patients. The framework can be beneficial for Tanzanian people and prove that hospitals can get better results and efficient care through an integrated and organized healthcare system. The figure below shows in detail how the framework should work. The common core component of the framework is an application suite, consisting of different operational application across Tanzania and integrated through a common operational database and this is important because it can ensure standard data and interfaces for clinicians and other users. In order to develop the proposed framework in Tanzania we introduced the following strategies.

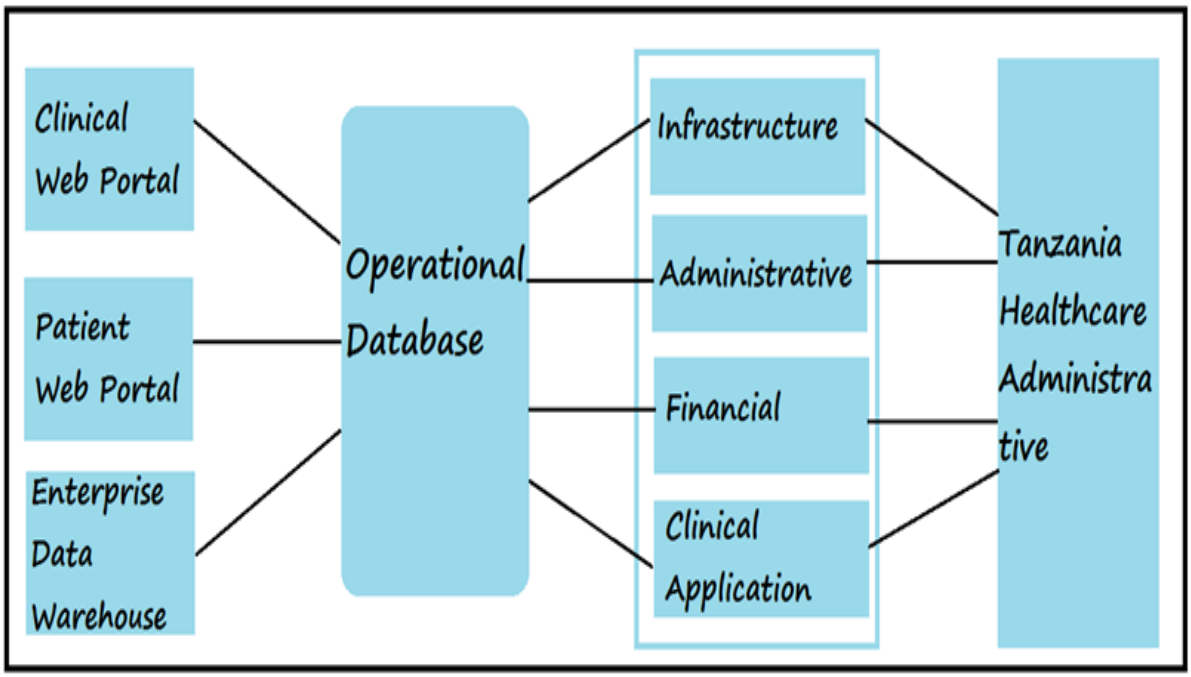

Figure 2: Proposed Framework for Tanzania Healthcare System

4.1.Clinical Data Exchange standards in Tanzania Healthcare System

The goal of clinical data exchange standards is to develop a comprehensive record of patients that will be available virtually anywhere in the country and accessible through any system. The lack of efficient data exchange is the major barrier of many healthcare systems across the globe, hence we should overcome this barrier in implementation. Once clinical data exchange have been implemented patients and drugs information should be available from one point to another. If this is not implemented, clinicians can face difficulties to exchange information with other clinicians across the country especially during disasters and emergency response situations. Also medical information cannot be readily available at the point of care.

4.2.Align proposed system with Clinical and Administrative Process

The Tanzania Healthcare proposed system may not improve patient care if the system is not aligned with clinical and operational processes. Clinical processes refers to the interdependent and collaborative activities that are performed to provide effective and efficient patient care, while administrative process refers to the interdependent and collaborative activities related to operational and financial matter pertinent to patient

care and organizational management. It is very important to take into consideration the alignment factor otherwise it can lead to system failure, another important factor is the role of the organization to use IT applications.

4.3.Web Based Interface for Tanzania Healthcare Administrative System

Advances in Internet and Internet based technologies have provided numerous opportunities not only in healthcare but in other sectors as well. Web based delivery is gaining momentum among other sectors, but in healthcare still more work is needed to be done. Hence, a common framework in healthcare needs to be designed and developed in order to boost the efficiency of the healthcare system in Tanzania. Also, the lack of security and privacy guidelines pertaining to patient information need to be structured. The web based system is the solution to provide robust and timely retrieval of patient data from any location across the country during disasters and emergencies. The system not only helps the clinician but also the patients and family members can be benefit as well.

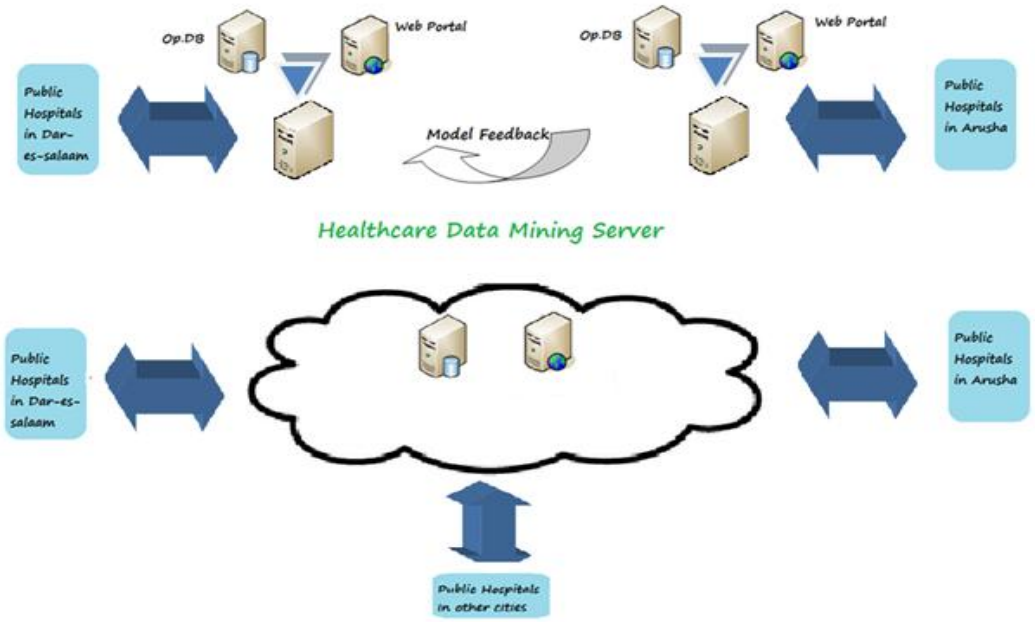

Figure 3: Proposed Architecture for Tanzania Healthcare System

4.4.Develop Enterprise Data Warehouse and Business Intelligence and integrate with proposed system

The goal of Enterprise Data Warehouse (EDW) is to capture and process important healthcare data where the decision making body wants to get the overview of the data and not the details of the data. The EDW architecture enables data from different operational systems across the country to be loaded through Extraction, Transformation and Loading (ETL) processes. Data Marts will be developed to structure the data from different subject areas of the warehouse such as outpatients encounters, inpatients encounters and pharmacy to enable clinicians and other users to access data through a common business intelligence and data analytic interface. The common interface powered by business intelligence and analytic tools presents the vast amount of patient data accumulated over a long period of time in aggregate fashion to understand long term pattern, efficiency and effectiveness of a certain procedure or medication. This involves patients care in two ways which clinicians can make better decisions and the data from EDW can be used in medical research.

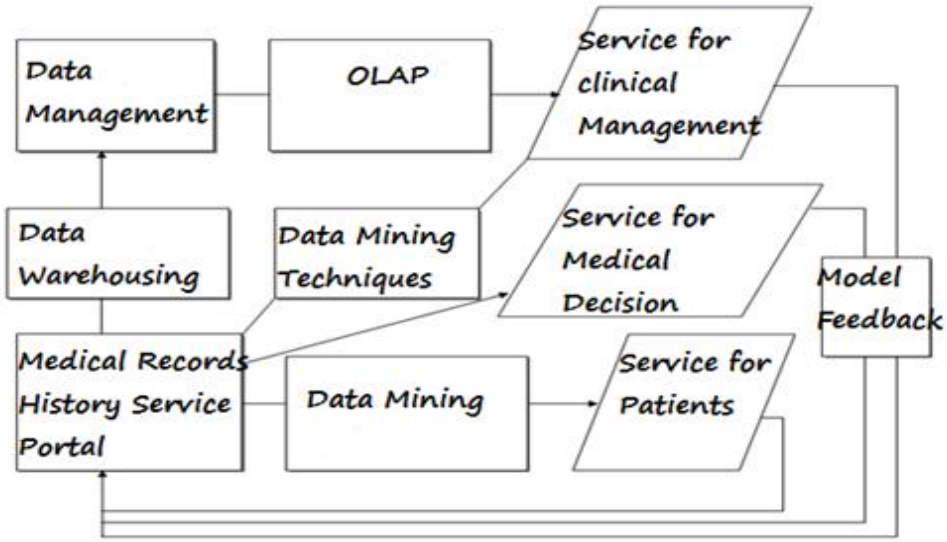

Figure 4: Proposed EDW Module for Tanzania Healthcare Administrative

Tanzania Healthcare Administrative will propose the plan to develop an EDW infrastructure created significant enterprise synergies, economies of scale and enabled the following:

- National Data Marts such as Lab, Pharmacy, Trauma, Pathology, Radiology, Primary care, Oncology, Administrative (workload, cost, demographic, utilization), Access management and Quality management
- Successful regional data warehouse by supplying standardized and cleansed data and by sharing best practice and knowledge
- Data/text mining, discovery and exploration for research and clinical purposes
- Enhanced national level registries such as HIV, TB, Malaria, Diabetes and cancer to support national effort and achieve better interoperability with partners such as Tanzania Ministry of Health and National Institute of Medical Research (NIMR) and other healthcare related NGOs.
- Feedback to other operational systems integrating analytic information to operational decision making.

4.5.Provide Decision Support Capability through Tanzania Healthcare proposed system

One of the major healthcare initiatives of the Tanzania Ministry of Health is to accelerate the diffusion and dissemination of clinical research data for policy makers, sponsors, researchers and medical community at large. Research findings and medical discoveries must be converted into useful products and service for physician, patients and healthcare providers. Clinical Decision Support System (DSS) is a very important component to enable this and can also substantially reduce the time of submission of higher quality research to Institute of Medical Research (NIMR). Eventually an interoperable network of Tanzania Health Administrative system is necessary to accelerate the process of transforming research into practice by integrating into national and regional database of clinical DSS and thereby delivering up to date knowledge of clinician at point of care. Clinical DSS can help reduce the risk to public health from dangers such as communicable diseases, hazardous or unsafe foods and other catastrophes by disseminating critical information at the right time. In emergency it is absolutely necessary to alert both clinician and consumers quickly. DSS can be updated and integrated with systems from hospital, medical centers and public health agencies, thereby giving public health professionals all the necessary information regarding the medical health to react early.

V. BENEFITS

5.1.Improve healthcare management efficiency

Data mining in healthcare can be able to identify and track patients in order to design appropriate methods and algorithms as a means of lowering number of cases of diseases as well as patients medical claims. Using web portal patients can search for their medical related problems, hence improving their knowledge regarding their health related issues, and also can have one on one discussion with their physicians.

5.2.Better Patients Physician relationship

Patients-physician relationship is an important aspect within the healthcare sector. By understanding patients " needs and wants we can significantly improve their level of satisfaction. Hence data mining can help to find the hidden pattern for patients" needs and wants from their healthcare providers.

5.3.Decreased Insurance Frauds

Fraud related issues recently occurred in the National Health Insurance (NHIF) of Tanzania and this sector is very new here. Data mining can be of a big help in the healthcare insurance frauds. It has the ability to detect and identify fraud based on the situation by finding the hidden pattern. Data mining can be able to find any abnormal behavior related to fraud and medical claims.

VI. DISCUSSION

One of the major challenges in the adoption of healthcare in developing countries is lack of support from major stakeholders, lack of patient unique identification, lack of funds, lack manpower, confidentiality and security. The Institute of Medicine (IOM) in 1999 shocked the nation by reporting that as much as 98,000 people die in hospitals every year due to medical errors. These errors are also said to cost hospitals as much as $29 billion every year. Of the many reasons identified for the medical errors, one critical reason is the decentralized and fragmented nature of information related to patients, drugs, procedures and medical processes. IOM also reported that about three out of four errors could have been eliminated by better healthcare system to make drugs and patients information readily available when needed [16]. We argue that all these challenges can be overcome if we have trust and strategies to implement such a system. The government of Tanzania is in the process of implementing the National identification card which is the unique ID for every Tanzania Citizen which is ran under National Identification Authority (NIDA). This is a good start but still the implementation process is still slow, and so we propose that the process should move a little bit faster and in actionable manner. The government of Tanzania is making a lot of effort to overcome the manpower in healthcare by establishing healthcare institutions, encouraging students to take science subjects, and also organizing scientific conferences like the National Human resource for health which is taking place for the first time in Tanzania with a major agenda to discuss issues relating to manpower in health care and how to go about it. The conference is organized by Benjamin Mkapa foundation, and we propose this type of conferences to take place more often. Also we propose to use free open source software due to the lack of fund and the government to make more efforts in training healthcare experts.

VII. CONCLUSION

Healthcare is one of the major sectors which can highly benefit from the implementation and use of information system. We have provided an overview of applications of data mining in infrastructure, administrative, financial and clinical Health care system. We proposed a best fit data mining framework that can greatly improve the healthcare sector in Tanzania. We discussed in detail how clinical data warehousing together with data mining can improve healthcare system in Tanzania. The proposed framework presented here can greatly benefit the healthcare sector by improving the quality of patients care, reducing medical cost, reducing time to wait for medical treatment and improve patient physician relationships. Despite those benefits we still have big challenges ahead of us such as high cost of implementation, lack of support from important stakeholders, lack of patient unique identifier, lack of healthcare policies, lack of man-power and lack of privacy, confidentiality and security concerns.

ACKNOWLEDGEMENTS

I would like to thank almighty and my family for their constant support during this work. I would also like to thank my supervisors Dr. Anael Sam from Nelson Mandela African Institute of Science and Technology and Dr. Muhammad Abulaish from Jamia Milia Islamia and last Dr. Yaw Nkansya-Gyekye from Nelson Mandela African Institution of Science and Technology for their guidance and valuable support during this work.

REFERENCES

[1] Nemati, H.R., & Barko, C.D. (2002). 'Enhancing enterprise decision through organizational data mining'. The Journal of Computer Information Systems, 42(4), 21-28.
[2] Lampe, J. C., & Garcia, A. (2004). 'Data Mining: An In-Depth Look'. *Internal Auditing*, 19(2), 4-20.
[3] http://www.ikanow.com/blog/02/21/data-mining-for-healthcare--proven-remedy-for-an-ailing-industry
[4] Greens R., "Clinical Decision Support". Elsevier Inc., 2007.
[5] Chen Y., Henning Pedersen L., Wesley W. Chu, Olsen J., "Drug Exposure Side Effects from Mining Pregnancy Data". ACM SIGKDD Explorations Newsletter, 2007.

[6] Verma, R. Harper, J., "Life Cycle of a Data Warehousing Project in Healthcare". JOURNAL OF HEALTHCARE INFORMATION MANAGEMENT, VOL 15, PART 2, pages 107-118, 2001
[7] Kevin Campbell, N. Marcus Thygeson and Stuart Speedie. Exploration of Classification Techniques as a Treatment Decision Support Tool for Patients with Uterine Fibroids; Proceedings of International Workshop on Data Mining for HealthCare Management, PAKDD-2010.
[8] Laudon and Loudon,(2004), Data Mining techniques and customer behaviors guide in decision making
[9] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). 'The KDD process for extracting useful knowledge from volumes of data'. Association for Computing Machinery Communications of the ACM, 39(11), 27-34.
[10] Brabazon, T. (1997, Jun). Data mining: A new source of competitive advantage Accountancy Ireland, *29*, 30-31.
[11] Firestone, J. M. 'Data Mining and KDD: A Shifting Mosaic'. (1997, March 12, 1997). from www.dkms.com/papers/dmkdd.pdf (23 April, 2004)
[12] Berry, M. J. A., & Linoff, G. (1997). Data Mining Techniques For Marketing, Sales, and Customer Support. New York: John Wiley & Sons, Inc.
[13] Nemati, H. R., & Barko, C. D. (2002). 'Enhancing enterprise decisions through organizational data mining'. The Journal of Computer Information Systems, 42(4), 21-28.
[14] Fong, A. C. M., Hui, S.C., and Jha, G. (2002). 'Data Mining for Decision Support'. IEEE
[15] Wen, C. P. (2004). 'Hierarchical Analysis For Discovering Knowledge in Large Databases'. Information Systems Management, 81-88.
[16] Institute of Medicine. To err is human: Building a safer health system. In L. T. Cohn., J. M. Corrigan., and M. S. Donaldson (Eds.) Committee on the Quality of Health Care in America. National Academies Press, Washington D.C., 1999.
[17] http://www.mkapahivfoundation.org/index.php/en/
[18] veteran affairs transformed itself and what it means for the rest of us. Business Week (July 17, 2006), 50-56.
[19] Brown, S. A., Lincoln, M. J., Groen, P. J., and Kolodner R. M. VistA—U.S. Department of Veterans Affairs national-scale HIS. International Journal of Medical Informatics 69, (2003), 135-156.
[20] Carpenter, D. Spending spree. *Hospitals and Health Networks 79*, 5 (2005), 4-7.
[21] Hillestad, R., Bigelow, J., Bower, A., Girosi, F., Meili, R., Scoville, R., and Taylor, R. Can electronic medical record systems transform health care?: Potential health benefits, savings, and costs. Health Affairs 24, 5 (2005), 1103-1117.
[22] Frequently asked questions. URL: http://www.ideatel.com/info.html