A Flexible OLAP Query Model for a Telemedicine System

Alexander Salamanca, Francisca Rojas, Tanya Romero, Eduardo Romero
Grupo de Investigación BioIngenium
Centro de Telemedicina
Universidad Nacional de Colombia
Bogotá, Colombia
[pasalamancar,frojass,tkromeroa,edromero]@unal.edu.co

ABSTRACT

This article presents the design of a information system layer to support analytical telemedicine services. We introduce general considerations on the chain of telemedicine value, as well as a query tool that allows flexible and efficient querying of relevant information about performance, scope, effectiveness, and measurements of disease outcomes related to a risk populations with a telemedicine data warehouse. For decision-makers, this information is vital to identify both the opportunities and the emerging problems.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

Keywords

olap, data warehouse, HL7, Information systems, telemedicine

1. INTRODUCTION

Telemedicine can be considered as a tool which allows access of highly specialized medical services to underserved regions. Specialties such as dermatology, radiology or infection diseases are nowadays available for such regions thanks to communication networks, capable to support great information traffic. Currently, the Telemedicine Centre[1] is an institution offering telemedicine services with coverage on several regions of Colombia. Considering the chain of value of the Centre in terms of social, economic and research impact generated in the service providing, it is needed to rely on tools and processes which provides support to the information and knowledge management.

In this work, we present the design and development of an analytic layer over our existent transactional telemedicine system[16], whereby effective support to decision makers on health field can be provided by means of a flexible query system which retrieve information stored in the data warehouse.

[1] www.telemedicina.unal.edu.co

EATIS '08, September 10-12, 2008, Aracaju, Brazil

Also, we complement the data warehouse with a query matching module in such way that the decision makers can have easy access to a lot of information in order to handle it and interpret it efficiently. The query matching system allows to make queries on keyword search text, as they are automatically matched to most relevant predesigned multidimensional queries, and then give the possibility of modifying and crossing the corresponding variables of interest.

This paper is organized as follows: in section 2, we make a brief presentation of datawarehousing and multidimensional queries. In the section 3, it is explained the architecture proposed. In section 4, the methodology used for this work. In section 5, we present related work. Finally, in the last section we present conclusions and future work.

2. BACKGROUND

According to [11], a data warehouse is the combination of the four main components: *data staging area* where source data are cleaned, transformed, combined, and prepared for its use; *data access tools*: where data are organized, stored, and made available for direct querying; *data presentation area*, where another analytical applications can be used. A data warehouse must make the information easily accessible, it must present that information consistently and it must serve as the basement for improved decision making. The above requirements impose that the data warehouse must support the needs of organization goals. In this work we design components for access and presentation of data as it is highlighted by the segmented line in the figure 1.

In this sense, a data warehouse is a repository into which all data relevant to the management of an organization are placed and from which it emerge the information and knowledge needed to effectively manage the organization [19]. A key role of the data warehouse is to provide compelling business intelligence to the decision-maker, facilitating an understanding of business problems, opportunities, and performance. It must incorporate internal and external knowledge acquired over time and adapt it to current business conditions [14].

The primary consumers of the data warehouse contents and business intelligence applications are decision makers and knowledge workers within the organization. Often, performance of these end users is measured by how effectively they are able to use available data to make good decisions. Many of the end users are not technically oriented and need

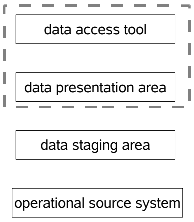

Figure 1: Data warehouse components. The scope of this work is highlighted with the segmented line

a significant amount of support to use a data warehouse effectively [8].

3. ARCHITECTURE

In this section we tackle with the architecture of the proposed system. It should be strengthen out that we mainly focus on the theoretical framework on which the proposed model rather than implementation issues.

3.1 Telemedicine System Information Database Architecture

The Telemedicine Centre platform aims to satisfy health care needs integrating modules for telemedicine activities, where medical processes are modeled following the HL7 Reference Information Model. HL7 is a very adaptable exchange message model which allows to easily communicate between very different activities. For instance, straightforward communication and support between medical and administrative processes. This has allowed easy inclusion of many specialities such as dermatology, radiology, cardiology, pathology and infection diseases, among others with no need of changing nothing in the underlying database structure. The system is able to implement security mechanisms such as Digital Signature and Zero Knowledge Proof for authentication or a telecare module which measures patient's vital signs in realtime.

The figure 2 shows the system layers, the Business Logic layer includes a Service Oriented Architecture (SOA), where thanks to the component granularity shows a low coupling is achieved. In this design, client accesses services through interfaces, whereby implementation of services is independent of the specific service definition.

Therefore, required changes in the GUI, only affect the Web application, not the nuclear components. Likewise, changes in the WebService implementation or in any EJB (Enterprise Java Bean) need not affect the GUI. Also, in the diagram 3 is presented the separation between several modules that compounds the entire telemedicine system.

3.2 The proposed model: OLAP applied on structured data base under HL7

HL7 provides an appropriate methodology for structuring the database under high quality standards in health and therefore the resulting information allows to support multiple different tasks. This model allows thus information to be exchanged between software applications, while conserving information semantics. A structured database under the

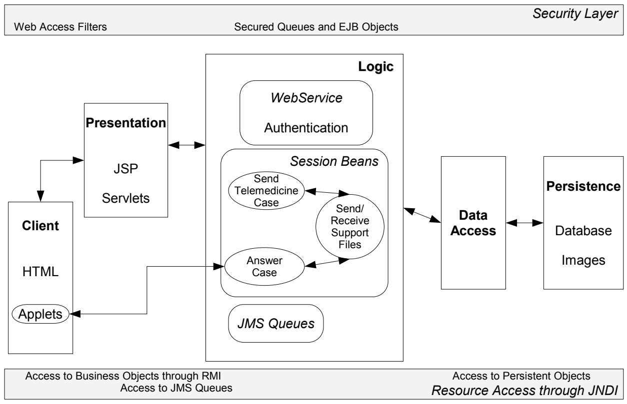

Figure 2: Layered Architecture Diagram

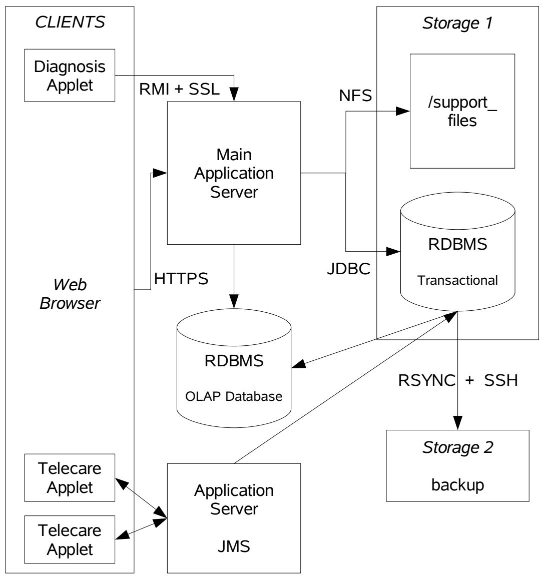

Figure 3: Architecture Diagram

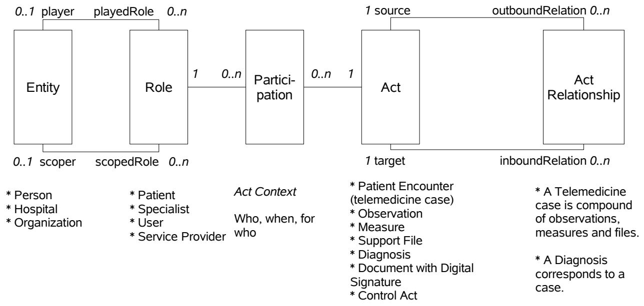

Figure 4: Reference model adapted to Telemedicine Information System

HL7 standard ensures the integrity of health data, but this model is not devised for efficient data consultation so that an additional module must be included. A data warehousing module can then take advantage of this large number of data to generate structured consultations (OLAP) that will be used by personnel decision-making in the health centre.

The combination of perfectly structured data (HL7) with efficient consultation (Data warehousing), results in a powerful system that provides high-quality consultations on health field.

In addition to the models system, graphics interface plays an important role, beyond being a beautiful presentation of consultation system; the interface is what allows so graphically contrast the dimensions of the consultations.

3.3 The HL7 Reference Model

The transactional component design is based on Health Level 7 (HL7). HL7 is an organization who is responsible for developing standards in health care and who is accredited by the ANSI (American National Standards Institute). HL7 efforts are focused on administrative and clinical data management. An important initiative proposed by this organisation is called the Reference Information Model (RIM), an abstract model for medical information management and whose backbone classes comprise act, entity, role, participation, act relationship and role link. Each backbone class inherits to many subclasses which provide more specificity to the data model. Also, this structure permits to add new subclasses, based on the RIM's classes, whereby it is quite easy any adaptation of the reference model to the particular needs of health entities. Overall, classes are related with each others such as next figure 4 illustrates.

3.4 OLAP Query Model

OLAP or Online Analytical Processing is a business intelligence approach. OLAP provides answers from multidimensional queries on a database. According to literature [11], it is established three kinds of data warehouse applications: information processing (supports querying, basic statistical analysis, and reporting using crosstabs, tables, charts and graphs), analytical processing (multidimensional analysis of data warehouse data and supports basic OLAP operations, slice-dice, drilling, pivoting) and Data mining. (Knowledge discovery from hidden patterns supports associations, constructing analytical models, performing classification and prediction, and presenting the mining results using visualization tools. OLAP focuses on interactive data aggregation tools; data mining emphasizes more automation, deeper analysis).

4. METHODOLOGY

To design and develop our analytical query system we propose: first, we make general considerations of the core processes of telemedicine services, then we design the data warehouse. Second, we introduce the architecture for the data query processing interface that was built; specifically, we enhance it with a novel approach to match an information need requested by an user with a predesigned set of multidimensional MDX queries. Finally, the prototype user interface is developed.

4.1 Datawarehouse

We are going to approach the design of a dimensional database by considering the four steps as proposed in [11]. 1) Select the business process to model. 2) Declare the grain of the business process. 3) Choose dimensions. 4) Identify facts.

Identifying the chain of value

First at all, we establish the business process of interest centered in our main chain of value. For our purposes, we can define a business process as a collection of linked activities that consume inputs, add value, and produce an output of value to an internal or external customer. In order to find our target processes, we need to investigate which have the utter impact on this organization, which add the largest value for the end customer, which are the most expensive to operate, and which have the more perceived problems such as defined in [6]. In this sense, three main processes of multiple levels were chosen: tactic, strategic and operative. These will be characterized as follows.

Sending and receiving cases of telemedicine. *Goal*: to bring specialized care to the patient to remote populations. *Roles*: internist, especialist, nurse, patient, support technician. *Activities*: to fill out the clinic history of the patient, to attach any diagnostic images or support files, to send the telemedicine case; check out the case and submit a diagnostic; real-time patient monitoring of signals. *Input*: clinic history of the patient, with images and another support files. *Output*: diagnosis, recommendations and treatment. *Resources*: patient, physician at the remote station, specialist, nurse, remote workstation (equipped with a PC, a scanner, and a digital camera), reference centre workstation (equipped with a PC, with high res monochromatic screen), a server and a datacenter.

Epidemiological surveillance. *Goal*: to identify risk factors and health trends of the attended population. *Activities*: to establish comparison criteria, benchmarking and formulation of indicators and strategies. *Input*: data warehouse. *Output*: strategies. *Resources*: information contained in the data warehouse and government databases.

Evaluation of operating performance in terms of market share and coverage. *Goal*: to improve the levels and quality of coverage. *Activities*: establishment of criteria, determination of target population, formulate the balance

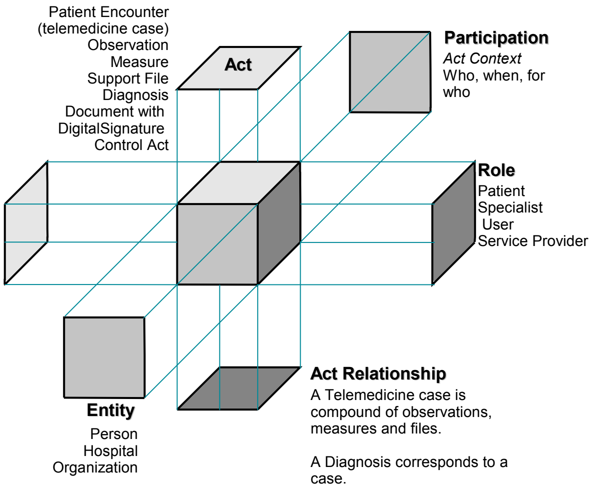

Figure 5: Our fact table and their dimensions

scorecard. *Input*: data warehouse. *Output*: strategies. *Resources*: information contained in the data warehouse and government databases.

Dimensions and attributes

For the dimensional model we have selected the dimensions depicted in figure 5. Considering our base platform, it can be extended easily to incorporate dimensions and their attributes like as follows. *Diagnosed Disease*: code, name. *Time of Day*: hour, working hour. *Speciality*: code, name. *Remote Station*: code, name, zone, department, contracting, type of contract. *Calendar Date*: code, date, day of week, week number, month, season, holiday indicator. *Specialist*: code, id, speciality name, age, genre. *Referral doctor*: code, id, age, genre. *Patient*: code, id, age, genre, activity, place of birth, department of birth, place of procedence, department of procedence. *Reference Centre*: code, name, zone, category. *Contract*: code, amount, start date, end date, type. *Administrative Resources*. This is achieved so as to allow maximum flexibility and dimensionality with the most atomic data. In such way we can satisfy information predetermined needs or facilitate the construction of unexpected reports for measuring the previous processes.

Facts and measures

The measures in our fact table are summarized here: number of cases, number of auxiliary files, presumptive diagnostic, and definitive diagnostic. As we can see, we have two types of measurements of interest, namely: numeric and text. In this paper we only tackle with numeric measurements.

Implementation

For implementation of the data warehouse and integration with the current telemedicine system we used the Mondrian OLAP Server[2]. It enables to interactively analyze very large datasets stored in SQL databases. As usual in datawarehousing, we have separate databases with their respective data. An automatic daily update has been set up from the

[2]mondrian.pentaho.org

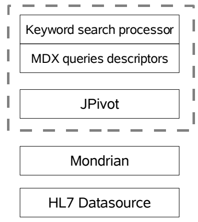

Figure 6: Components of our Query Model

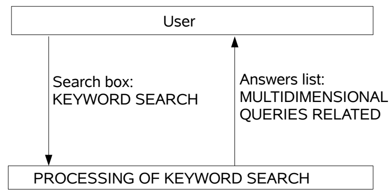

Figure 7: Query Interaction Diagram

transactional telemedicine database. In order to access to this multidimensional model the MDX language is used. In figure 6 its shown how the components proposed fits the considerations at the beginning of the paper.

4.2 Query Processing Tool

In order to establish a set of predetermined and pertinent needs of information, a series of information needs solvable by MDX queries in the datawarehouse were set by the Centre management team. Also was requested a group of textual descriptors for every one. These are used as descriptors for the retrieval process with keyword search queries requested by information users.

4.3 Prototype User Interface

The system allows the user to perform a query in his normal language, typing his question into a search box. The mechanism used to search for a suitable query given the information need involves a very simple workflow as can be seen in the figure 7, where the user can submit a keyword text query to find a relevant set of predesigned MDX queries.

Each multidimensional query stored in the database is associated with a text descriptor (here that will be called "document"), so we can construct a vectorial representation of the collection using the traditional TF-IDF representation [15] after some preprocessing steps as depicted in the figure 8. There is a vector for each document, and the same preprocessing can be applied for query text. So we can calculate a measure of similarity between the query vector and documents vectors using the cosine distance as common in the retrieval literature [13]. The answers are ranked in order of relevance to the query text. For the implementation of this model, we used Lucene, a powerful open-source indexer[3]. For a given query, Lucene returns a ranked list of documents, based on an optimized formula combining a mixture of vector space and boolean similarity.

[3]lucene.apache.org

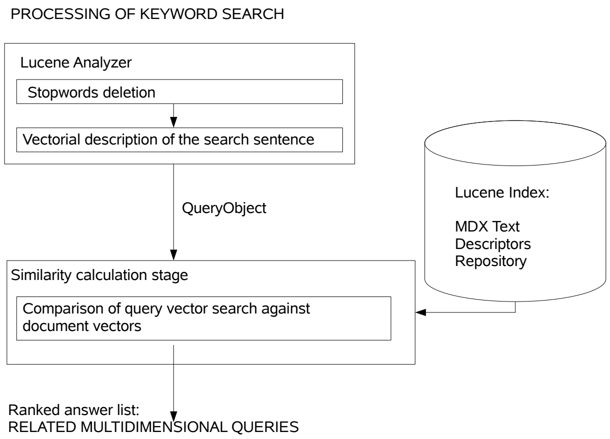

Figure 8: User Query Processing

Figure 9: Prototype Interface for query submission

Query matching
The user interface is designed, having in mind, that is a very important part for the user, because indeed, thats the system for the user. In this system the data is completely explorable and can be visualized in an interactive manner, due that the display mode of the results can be changed through filters and views, in an asynchronous form respect to the query on the database.

Data interaction
Initially, the user finds a search box that receive keyword text queries to select views of data, as well as possible suggestions to user's query. This leads to expected outcome easier and let the user choose among potential MDX queries to exploration (see figure 10). The interface design (figure 9) shows great flexibility of queries that can be submitted to the analytical layer and it also supports the ability to display the results graphically so the user can cross more variables and interact upon the initial search.

5. PRELIMINARY EVALUATION

Currently, more than 30 hospitals of first and second level complexity use the telemedicine system for remission of cases, and more than 20 specialists, work for answering and solving the cases. These hospitals are health centers for their respective underserved regions and their action radius covers populations which vary between 200.000 and one million

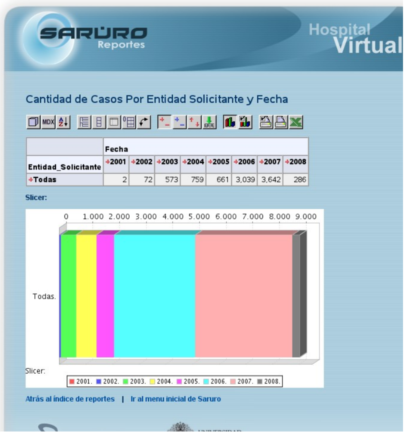

Figure 10: A current interaction in a MDX Query

inhabitants. After 15 months of service, the system processed 3.584 actual telemedicine cases from 31 remote stations with an average response time of 1,6 days, showing flexibility, security and scalability. Over this, our analytic system has been performing since August of 2007, where it has been possible conducting reports for the hospitals, policy makers and for government entities, where reports were built in a timely manner with a set of nearly 30 predefined MDX queries.

6. RELATED WORK

In the literature, there are some published implementations of data warehousing in the health field [3, 17, 10, 1, 18, 9, 12, 4, 7, 5, 2], but most of these are limited to traditional models of querying and reporting.

As part of a larger project to design the necessary architecture for multi-institutional sharing of disparate biomedical data, [12] explored the potential of the HL7 Reference Information Model (RIM) for representing the data stored in a local academic clinical data warehouse and he describes his initial efforts at mapping clinical concepts from a relational data warehouse to the HL7 RIM. A telemedicine system that allows the chronic HIV/AIDS patient to improve his/her self-care and to be remotely followed-up by the healthcare professionals is presented by [4]. Otherwise, [7] developed a Doctor Workstation System (DWS). This DWS implements the conformity of the medical information systems by applying semantic web and web service techniques. On the other hand, [5] describes the usage of a computerized patient information system,that allows for timely and concise extraction of patient data from the computerized medical record.

Management of healthcare is seen to involves data at every encounter of each patient with any provider, pharmacy, payer, or government agency. These data and their analysis are massive on multiple dimensions: of patient-encounter records; of variables (administrative, diagnostic and procedural) and their derived indicators; of the related clinical knowledge resources; of the clinical and administrative issues to be addressed; and of the diversity of the audience for the analysis [9].

Regarding our approach to querying, this is most similar to the employed in other fields like the recently released by United Nations Organization: "UNdata, A World of Information"[4]. Its design also allows a user to access a large number of UN databases either by browsing the data series or through a keyword search.

In this sense, our work is original since we tackled the user interaction with the data warehouse in an easy way for the users. We didn't find similar models reported in the literature, then there's no comparison with others systems based on HL7 that allows analyze the queries and match it with descriptors and that allows .

7. CONCLUSIONS AND FUTURE WORK

This paper presents a flexible and interactive analytical layer on the telemedicine platform. It can be hypothesized that system performance depends on the quality of descriptors suggested by experts, but this needs to be tested furthermore. Also, is needed to consider aggregations over textual measures on the table of facts, particularly, because these can be very useful to gain insights about the medical prescriptions.

We have foreseen to evaluate the system with a extensive test, due to the preliminary evaluation only considered staff user satisfaction with no exhaustive testing and also we no have in account the performance of the information retrieval method. This design is geared to address the needs of users, and acceptance of the preliminary prototype has been very positive in our users.

8. REFERENCES

[1] S. S. R. Abidi. Knowledge management in healthcare: towards 'knowledge-driven' decision-support services. *International Journal of Medical Informatics*, 63:5–18, 2001.

[2] B. Blazona and M. Koncar. Hl7 and dicom based integration of radiology departments with healthcare enterprise information systems. *International Journal of Medical Informatics*, 76S:S425–S432, 2007.

[3] J. L. Breault. Data mining a diabetic data warehouse. *Artificial Intelligence in Medicine*, 26:37–54, 2002.

[4] C. Caceres, E. Gomez, F. Garcia, J. Gatell, and F. Pozoa. An integral care telemedicine system for hiv/aids patients. *International Journal of Medical Informatics*, 75:638–642, 2006.

[5] J. Camarro. Collecting patient research data using a computerized patient information system. *Nephrology Nursing Journal*, 35, 2008.

[6] M. Cumberlidge. *Business Process Management with JBoss jBPM.* Packt Publishing, 2007.

[7] M. Fang, Y. Zhang, H. Zhao, and N. Wu. Integration of hospital information systems in china. *New Mathematics and Natural Computation*, 3(1):135 – 151, 2007.

[8] N. Foshay. Does data warehouse end-user metadata add value? *Communications of the ACM*, 50, No. 11:70–77, 2007.

[9] C. R. Goodall. Data mining of massive datasets in healthcare. *Journal of Computational and Graphical Statistics*, Vol. 8, No. 3:620–634, 1999.

[10] A. Granta. Integrating feedback from a clinical data warehouse into practice organisation. *International Journal of Medical Informatics*, 75:232–239, 2006.

[11] R. Kimball and M. Ross. *The Data Warehouse Toolkit.* Wiley Computer Publishing, 2002.

[12] J. Lyman, K. Scully, S. Tropello, J. Boyd, J. Dalton, S. Pelletier, and C. Egyhazy. Mapping from a clinical data warehouse to the hl7 reference information model. *AMIA Annual Symposium Proceedings*, 920, 2003.

[13] C. D. Manning, P. Raghavan, and H. Schtze. *An Introduction to Information Retrieval.* Cambridge U.P., 2007.

[14] S. T. March. Integrated decision support systems: A data warehousing perspective. *Decision Support Systems*, 43:1031–1043, 2007.

[15] G. Salton. *Dynamic information and library processing.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1975.

[16] C. Sanchez, A. Rueda, and E. Romero. A granular prototype for telemedicine based on hl7 information model. In *Congreso Colombiano de Computación - CCC 2007*, 2007.

[17] M. Taylor, J. Saltz, and J. H. N. and. Design of an integrated clinical data warehouse, 2004.

[18] Terziyan. The decision support system for telemedicine based on multiple expertise. *International Journal of Medical Informatics*, 49:217–229, 1998.

[19] H. Watson. Developments in data warehousing. *Communications of the AIS*, 8, 2001.

[4] data.un.org