Health Data Management in the Medical Arena

ANTÓNIO ABELHA, JOSÉ MACHADO, VICTOR ALVES and JOSÉ NEVES
Departamento de Informática
Universidade do Minho
Braga, Portugal
{abelha,jmac,valves,jneves}@di.uminho.pt}
http://www.di.uminho.pt/{~abelha,~jmac,~valves,~jneves}

Abstract: - In this paper it is presented an Agency for Integration, Archive and Diffusion of Medical Information (AIDA), which configures a data warehouse, developed using Multi-Agent technology. AIDA is like a symbiont, with a close association with core applications present at any health care facility, such as the Picture Archive Communication System, the Radiological Information System or the Electronic Medical Record Information System. Multi-Agent Systems also configure a new methodology for problem solving.

Key-Words: - Multi-Agent Systems, Data Warehousing, Medical Applications.

1 Introduction

Medical Information Systems (MIS) are seen as a way of optimizing the use of the existing health care infrastructure, without resorting to new and costly infrastructure (re)construction. The qualitative design of such an environment requires a basic understanding of patient and physicians related characteristics and capabilities. Patient care, patient education, medical education, and clinical research need to be considered to meet the basic requirements on the level of desirable services, determined on the basis of the patient's length of stay; i.e. used for modeling the significant entities of such a world. The aim is to extract conclusions for the level of services provided to the users.

Medicine has been for some years a very attractive domain for Computer Science (CS) researchers, in general. There is a great potential for information automation, and a lot remains to be done. Medical Informatics (MI) is indeed becoming an issue of study in which Medicine, Artificial Intelligence (AI), Mathematical Logic and Computing may overlap. Another reason for this increasing interest is costs. Today's strained health-care economics makes it necessary for expensive resources to be efficiently used, and requires a balanced management. CS researchers have long used Medicine to elaborate on their own work. The field is probably one of the most knowledge intensive ones, loaded with human reasoning, with most of the procedure relying exclusively on clinical experts. This makes health-care a perfect target for CS, since conventional systems are naturally bounded by their lack of rich knowledge representation and proof schemes. Medicine allows for the testing and exposing of new ideas and techniques by CS scientists, which creates some complicity between the two communities.

What is the best approach to help physicians and patients? It is at this point that the different disciplines overlap. A system intended to be a support tool for clinicians must have its focus on knowledge representation and reasoning schemes. Such system should be able to explain and justify its conclusions. On the other hand it has to be flexible enough to allow for simulation, training of novices and maintenance [1].

2 An Agency for Integration, Archive and Diffusion of Medical Information

Information sources in a healthcare unit are distributed, heterogeneous, large and complex; integrate medical equipments that speak different languages; are built around information systems customized by several companies using different operating systems, languages, applications and hardware. Communications are sometimes limited by old infrastructures and new projects collide with financial restrictions and bureaucratic delays. The homogeneity of clinical, medical and administrative systems is not possible due to financial and technical restrictions, as well as functional needs. The solution is to integrate, diffuse and archive this information under a dynamic framework, in order to share this knowledge with every information system that needs it. AIDA – Agency for Integration, Diffusion and Archive of Medical Information [2] - is an agency that provides intelligent electronic workers, here called pro-active agents, and in charge of tasks such as communicating with the heterogeneous systems, sending and receiving information (e.g., medical or clinical reports, images, collections of data, prescriptions), managing and saving the information and answering to information requests, with the

necessary resources to their correct and in time accomplishment (Figure 1). AIDA also supports Web based services to facilitate the direct access to the information and communication facilities set by the humans, i.e. AIDA's construction follows the acceptance of simplicity, the conference of the achievement of common goals and the addressing of responsibilities; the main goals are to integrate, diffuse and archive large sets of information from heterogeneous sources (departments, services, units, computers, medical equipments); AIDA also provides tools in order to implement communication with human agents based on web based services. Under these presuppositions, a Healthcare Information System (HIS) will be addressed in terms of:

- The Administrative Information System (AIS), which intends to represent, manage and archive the administrative information during the episode (an episode is a collection of all the operations assigned to the patient since the beginning of the treatment until the end);
- The Medical Support Information System (MIS), which intends to represent, manage and archive the clinical information during the episode;
- The Nursing Support Information System (NIS), which intends to represent, manage and archive the nursing information during the episode;
- The Electronic Medical Record Information System (EMR);
- The Information Systems (DIS) of all the departments or services, in particular of the laboratories (Labs), Radiological Information System (RIS) and Medical Imaging (PACS - Picture Archive and Communication System), which deals with images in a DICOM format [3].

In order to implement this system, distributed by nature, intranet technology was used on the side of the end user. All the system interfaces are user-friendly, Web based and low cost. It uses freeware tools or software database packages which licenses belong to the Portuguese Health Ministry (e.g., Oracle software). The intranet was implemented using computers with LINUX as operating system. Communication among agents will be specified by defining the transportation (i.e., the process that allows for a message to be sent or received), the language (i.e., the meanings of a message), the ontology (i.e. the conversational structure) and the architecture (i.e. the system links according to given protocols).

A message is sent by an agent using XML [4]. According to the ontology, the message is processed, integrated and archived in large databases. The ontology is defined by the administrators and can be managed using web tools. The healthcare network is used for physical transportation.

XML is a meta-markup language and stands for Extensible Markup Language. It is a set of rules to

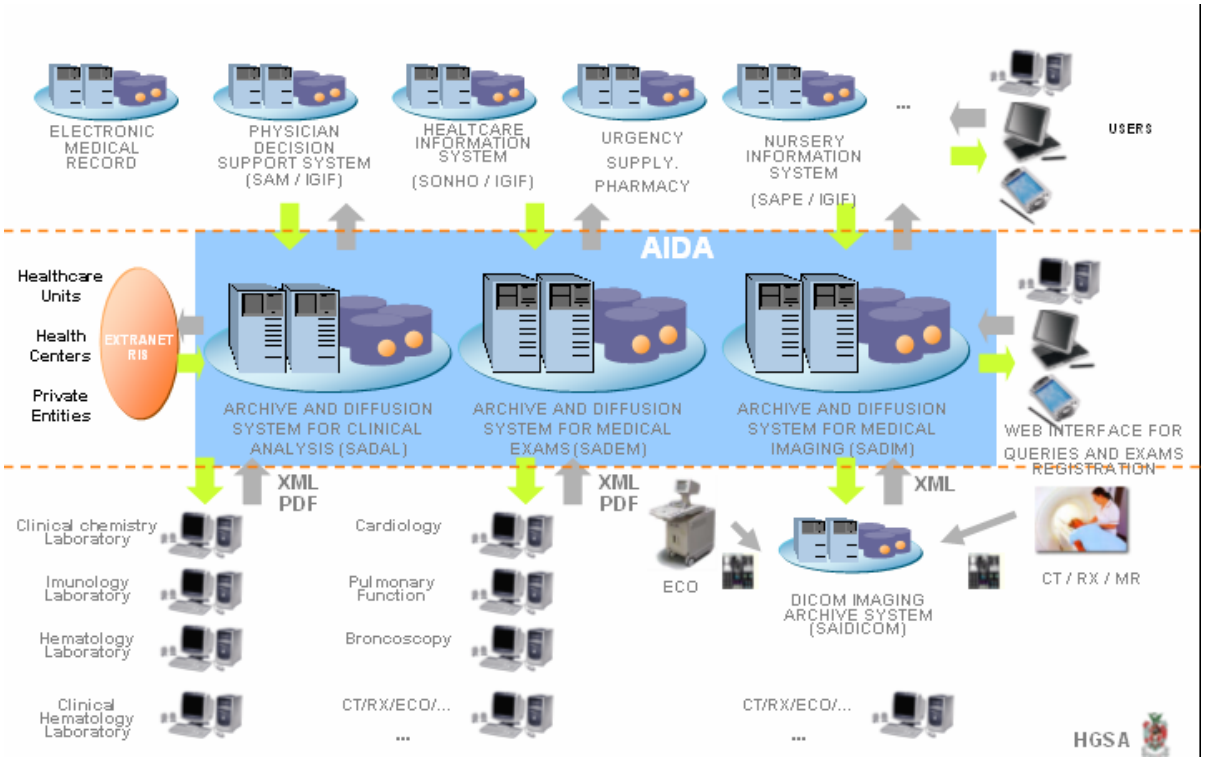

Figure 1- The AIDA modules

define semantics tags in documents, which are broken into parts. The language defines a syntax in which other field-specific markup languages can be written (e.g., HTML and TeX). HTML and TeX define a fixed set of tags that describe a fixed number of elements. In XML, data is not formatted to fit into paragraphs, text fields, list items, table cells or others general categories. Tags can be documented in a Document Type Definition (DTD); i.e., a vocabulary and syntax for some kinds of documents. XML has an excellent format for the interchange of data among different applications because it is free to use and easy for both human beings and computers to read and write. It is an obvious choice for exchange and share data, programs or languages between agents. XML also provides a client-side include mechanism that integrates data from multiple sources and displays it as a single document, what is very useful when working with large information repositories like RIS, PACS, EMR and AIDA. RDMBS applications (e.g., Oracle, MySQL, Microsoft) provide multiple programs to create, extract and parse XML documents from databases using Perl, Java, PHP or any other convenient language. AIDA uses XML documents to communicate between external data sources and operational databases. XML is also embedded in PHP programs that provide user interfaces.

The intelligence of the system as a whole arises from the interactions among all the system's components. The interfaces are based on Web-related front-ends that can be accessed using a standard Web browser, querying or managing the data warehouse [5][6]. Such an approach can provide decision support (e.g., with the radiologist conducting a form of dialogue with the technicians to extract knowledge and test hypothesis, using a strategy to compare a modality independent model with the image via an intermediate symbolic feature space).

3 Multi-Agent Systems

Multi-agent Systems (MAS) set a new paradigm in problem-solving via theorem proving, i.e. agent-based computing has been hailed as a significant break-through in problem solving and/or a new revolution in software development and analysis. Indeed, agents are the focus of intense interest on many sub-fields of Computer Science, being used in a wide variety of applications, ranging from small systems to large, open, complex and critical ones, i.e. agents are not only a very promising technology, but are emerging as a new way of thinking, a conceptual paradigm for analyzing problems and for designing systems, for dealing with complexity, distribution and interactivity, may be a new form of computing and intelligence.

To develop such systems, a standard specification method is required, and it is believed that one of the keywords for its wide acceptance is simplicity. Indeed, the use of intelligent agents to simulate human decision making in the medical arena offers the potential to set an appropriate software development and analysis practice and design methodology that do not distinguish between agent and human, until implementation. Being pushed in this way, the design process, the construction of such systems, in which humans and agents can be interchanged, is simplified, i.e. the modification and development in a constructive way, of multi-agent healthcare systems with a human-in-the-loop potential aptitude is becoming central in the process of agent-oriented software development and analysis. These systems have provided a clear means of monitoring the agent's behavior with significant impact in their process of knowledge acquisition and validation. MAS are a natural connection to intelligent systems evolution, being elements for task substitution or delegation, usually performed by human beings. However agent based systems have some restrictions, such as global system control and universal view absences, and some want of confidence and fear of competence delegation by human beings [7]. To delegate tasks, bilateral confidence relations have to be established. Organizations may also mature their experience relatively to the use of autonomous software components.

Agent theory in Artificial Intelligence deals with the behavior of autonomous and rational intelligent entities which are capable of executing actions in order to reach a given goal. Their behavior depends on intentional concepts like knowing, wanting, liking or believing, on actions that are executed depending on the states of the system, the level of the learning procedures and the context under which the conditions are tested. Indeed, actions are related to knowledge, the state of the world is perceived and plans for actions are formed. Actions may be periodic, spontaneous and may benefit the user. Agents may also take their own initiatives and may also be integrated on the World Wide Web (WWW) (Figure 2) (Figure 3). The problem is that the WWW is client-server oriented and WWW agents use peer-to-peer communication, so multiple messages from agents and collaborative tasks are not directly enabled. The WWW is oriented around structures

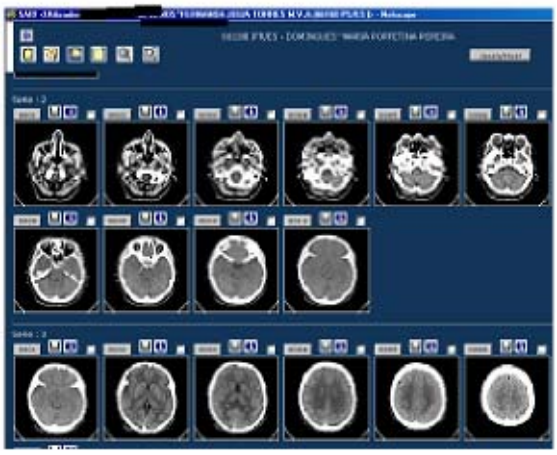

Figure 3- The PACS web interface

representing transport and display of information, in particular in Hyper Text Markup Language (HTML). With Extended Logic Programming coupled with Java based implementations of the Logic Programming paradigm, it is possible to conceive agents that send messages and interact with people or other agents through browsers, aiming at distributed interactive simulation environments, i.e. any complex software system can be viewed as a collection of independent cooperating units or entities, each of which implements a subset of an application's total functionality [8][9][10][11].

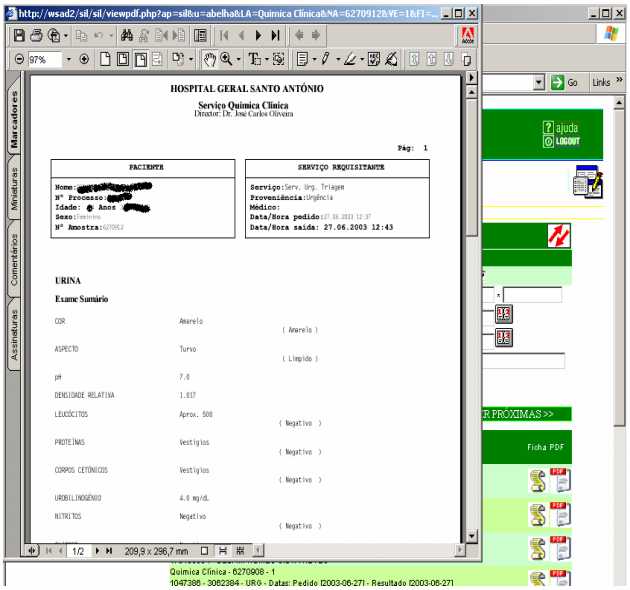

Figure 2- A Web Interface

4 Picture Archive Communication Systems

The computational architecture that supports *Medical Information* in the area of *Imagiology* is an example of an external data source for AIDA. To get the knowledge required to construct these system, one must be involved and familiarized with the DICOM standard, Computer Tomography (CT), Magnetic Resonance (MR), procedure documentation, debriefing of Neurologists. It also involves the regular participation of the authors in meetings at the premises of the healthcare unit [12][13]. A Architecture has then been envisaged to support the medical applications in terms of Intelligent Medical Information Systems, a form of a web spider of an intelligent information processing system, its major subsystems, their functional roles, and the flow of information and control among them, with adjustable autonomy. Indeed, many complex systems are made up of specialized subsystems which are understood as intelligent entities or agents that interact in flexible, goal-directed manners, and are understood as theories; i.e., the intelligence of such a system as a whole arises from the interactions among all the system's entities. There is also a purely communicative Multi-Agent System (MAS); i.e., there is no external environment influence and the agents communicate only through messages. A brief description of the different types of agents involved is given below:

- The Proxy Agents (PAs) which provide the bridges between the users and the system in terms of questions that may be formulated, explanations that may be required, decisions that may have to be taken and/or view of final results. The system's interfaces are based on Web-related front-ends using HTML pages, that can be accessed using a standard Web browser;
- The Decision Agents (DAs) which present mediating capabilities, act by accepting a task from the PAs. They may decompose it in sub-tasks, sending them to be processed in the CAs, and integrating the results (returned by the CAs);
- The Computing Agents (CAs) which accept requests for specific tasks from the DAs, returning the results;
- The Resource Agents (RAs) which present all the necessary information to their peers in order to allow them to access a specific data resource; and

- The Interaction and Explanation Agents (IEAs) which act based on argumentative proceedings which are fed with data and/or knowledge coming either from the PAs or the Das (note that the execution plans received from the DAs may be partial, so that only upon a completion of a task a trace can be compiled and an explanation can be delivered to the PAs and/or DAs).

Since their introduction, PACS have attracted a mixed reaction, mingling excitement at their potential with dissatisfaction with their various limitations. This has been true both in general AI, as in The Medicine field. Much work has therefore been concentrated on understanding these drawbacks, and developing problem solving methods and building strategies for such systems that may mitigate such faults. The overall trend has been away from seeing the process as one of encoding heuristics derived from an expert, towards modeling the domain on which the expertise operates. Further, it has become recognized that modeling requires as a precondition that the domain be conceptualized, and that the assumptions incorporated in the conceptualization must be made explicit.

However, if agent technologies are to be effective, design issues need careful consideration, i.e. how are MAS architectures to be designed for maxim effect?; can such architectures be used to support and enhance existing work practices?; what kind of information processing work and for what medicine's domains?; how can such systems be design to successfully complement people's existing practices and preferences? Indeed, what can be done with this conceptualization is now well understood in terms of its limitations, in the sense of the sort of medical domains to which it can be applied; its restrictions upon the duration or extent of an estate in the sense of the sort of system that will result from its application; and how to build and use PACS.

The system is characterized by the use of explicit anatomical models and for the visualization of the anatomical structures identified in the image segmentation. The anatomical model makes a major component of the system, and is organized in terms of a semantic network. The inference engine handles the decision making proceedings during the process of segmenting major anatomical landmarks.

The DICOM image server supports the medical interface – this window sets the via for the visualization and exploration of original DICOM data from CT, MR, and so on. It provides the user with interactive image visualization functions, like gray level windowing.

5 Data Warehouse for Healthcare

Relational Database Management Systems (RDBMS) are widely used by healthcare units for maintaining data that documents everyday operations [14]. The applications that update such operational data or transactions make typically small changes and a large number of transactions. RDBMS have been optimized to perform reliably and efficiently those operations, usually named online transaction processing (OLTP) applications (i.e., applications that support many users executing small transactions) [15]. Some RDBMS that support traditional SQL-style queries, were also designed to support Online Analytic Processing (OLAP) [16] queries efficiently. Such systems stand for RDBMS optimized for decision support applications. Here it is claimed that the natural solution in these cases is to create a centralized repository of all the data, i.e. a data warehouse managed by a MAS. It must support at the same time small transactions, executed by users, and very large transactions, executed by software agents, during the loading of data into the data warehouse, processes that can occur at times when online warehouse users are accessing the database, searching online updated records.

6 Radiological Information system

Radiological Information System (RIS) is a core application that is built around a MAS that receives information from heterogeneous sources (e.g., administrative applications, agendas, medical equipments, AIDA, PACS) and integrate it into the data warehouse in order to be managed by other external applications. Patient scheduling is made in real time and the integration with radiological equipments is made using DICOM worklist protocol through the MAS.

7 Electronic Medical Record

Electronic Medical Record (EMR) is also a core application which covers horizontally the health care unit and make possible a transverse analysis of medical records along the several services, units or treated pathologies. Beyond the organizational, functional, technical and scientific requisites, ones have to attend ethical and legal ones, as well as data quality, information security, access control and privacy. An Electronic Medical Record is an

assembly of standardized documents, ordered and concise, directed to the register of actions and medical procedures; a set of information compiled by physicians and others health professionals; a register of integral facts, containing all the information regarding patient health data; and a follow up of the risk values and clinical profile.

8 Conclusions

It is expected that AIDA, with a focus on integration, will ease the tension between the ideal and the pratical. Misaligned financial incentives and too few integrated, ease-to-use products, stymie health Information Technology adoption. The system is now at work at the Hospital Geral de Santo António, in Oporto, one of the two major healthcare facilities in the north of Portugal.

References:

[1] Abelha, A., *Sistemas Multiagente de Apoio a Trabalho Cooperativo em Unidades Hospitalares*, PhD Thesis, Departamento de Informática, Universidade do Minho, Braga, Portugal, 2004.

[2] Abelha A., Machado M., Santos M., Sollari A., Rua F., Paiva M. and Neves J., *Agency for Archive, Integration and Diffusion of Medical Inforation*, Proceeding of AIA'2003, IASTED Internation Conference on Artificial Intelligence, Benálmadena, Spain, 2003.

[3] Alves V., Neves J., Maia M., Nelas L., *A Computational Environment for Medical Diagnosis Support Systems*. ISMDA2001, Madrid, Spain, 2001.

[4] Harold E., XML Bible – Gold Edition, Hungry Minds Inc., NY, 2001.

[5] Chaudhuri S. and Dayal U., *An Overview of Data Warehousing and OLAP Technology*, SIGMOD record, 26(1), 1997.

[6] Kimball R., *The Data Warehouse Toolkit*, John Wiley and Sons, 1996.

[7] Heinze, C., Papasimeon, M., and Goss, S., *Specifying Agent Behavior with Use Cases, in Design and Applications of Intelligent Agents* – Proceedings of the Third Pacific Rim International Workshop on Multiagent Systems PRIMA 2000, eds. C. Zhang and V. Soo, 128-142.

[8] Neves, J., Machado, J., Analide, C., Novais, P., and Abelha, A. *Extended Logic Programming applied to the specification of Multi-agent Systems and their Computing Environment*, in Proceedings of the ICIPS'97 (Special Session on Intelligent Network Agent Systems), Beijing, China, 1997.

[9] Neves, J. *A Logic Interpreter to handle time and negation in logic data bases*, in Proceedings of the ACM 1984 Annual Conference, pages 50-54, San Francisco, California, USA, 1984.

[10] Alves, V., Neves, J., Maia, M., Nelas. L., *Computer Tomography based Diagnosis using Extended Logic Programming and Artificial Neural Networks*. Proceedings of the International NAISO Congress on Information Science Innovations ISI2001, Dubai, U.A.E., 2001.

[11] Traylor, B, and Gelfond, M. *Representing null values in logic programming*, in Proceedings of the 1993's International Logic Programming Conference Symposium (ILPS'93), pages 35-47, Vancouver, British Columbia, Canada, 1993.

[12] Neves, J., Alves V., Nelas L., Maia M., and Cruz R. *A Multi-Feature Image Classification System that Reduces the Cost-of-Quality Expenditure*, in Proceedings of the Second ICSC Symposium on Engineering of Intelligent Systems, Paisley, Scotland, UK, pages 594-554. 2000

[13] Looney K. and Theriault M., *DBA Handbook*, Oracle Press, Mc Graw Hill / Osbourne, 2002.

[14] Ramakrishman R. and Gehrke J., *Database Management System*, 3rd Edition, Mc Graw Hill Higher Education, 2003.

[15] Codd E., *Twelve Rules for On-line Analytic Processing*, Computerworld, 1995.

[16] Shoshami A., *OLAP and Statistical Databases: Similarities and Differences*, in ACM Symposium on Principles of Database Systems, 1997.