EMOSYS: A SYSTEM TO STUDY HEMATOLOGICAL DISEASES

Starita* A., Rossi* P. F. Caracciolo#, M. Petrini#
*Computer Science Dept., University of Pisa, Italy
# Department of Oncology, Transplants and New Technologies, University of Pisa, Italy

Prof. A. Starita, Dept. Of Computer Science, University of Pisa, Via Buonarroti 2, 56127 Pisa, Italy
Tel.: +390502212715, Fax: +390502212726, E-Mail: starita@di.unipi.it.

Abstract

This paper describes a web-based system named Emosys, developed both for managing and studying hematological diseases. The system is made of independent modules for the clinical management, for the collection of data and for research on these diseases. To achieve this goal two main cooperating agents have been implemented. The first allows sharing the clinical collection of data among doctors and clinical operators of the same department and communicating with remote clinical sites. The second one implements data mining techniques, based on neural networks, to confirm known results and/or to find interesting, unknown relationships among the data. The system is now under validation at the clinical site.

Key Words: Clinical Technology, Patient Monitoring, Intelligent decision support system, Data mining, Hematological Diseases.

1. Introduction

Nowadays the need of collecting and subsequently analyze a large amount of data has become of primarily importance in every research areas. This is especially true in a medical environment where it is very important to have at disposal an advanced and affordable software system to manage clinical sensitive data. In particular, this task is quite hard in a Hematological Unit (HU) due to the amount of patient’s data to manage and to the complexity and variety both of the diseases and related therapies. A web-based system, named Emosys, developed for studying hematological diseases, is here presented. Emosys has been designed in cooperation with the Hematological Unit of the Department of Oncology, Transplants and New Technologies in Medicine of the Santa Chiara Hospital in Pisa (Tuscany, Italy). Emosys implements a powerful web based, agreed EPR (Electronic Patient Record) that enables the hospitals, diffused on territory of Tuscany, to send their patient’s data to the Hematological Unit of Pisa that is the Regional Reference Centre. The network connectivity implements an ASP model so that the Central Unit administers the local main server and has full privileges over the entire data set, while each hospital can manage only the own data. This scenario allows building a structured, agreed database on which it is possible to apply classical and new Information Retrieval methods to confirm achieved results and/or to search for new knowledge. The Emosys system has been developed around three main objectives:

1. To collect reliable, homogeneous, agreed, data.
2. To design and implement a relational database of the patient’s clinical data.
3. To perform knowledge discovery in the realized database.

2. Materials and Methods

The Hematological Unit of the University of Pisa collects data about hematological diseases of subjects afferent to the S. Chiara Hospital of Pisa (the Regional Reference Centre of Tuscany) and it also receives data from external Hematological Units belonging to five hospitals diffused on the territory. This has required the design of a system that could allow sharing of data and experience with medical people from different and, some times, far sites. For this reason, a Web-based system has been developed so that the Unit of Pisa can control the quality of data and their correspondence to agreed protocols. Together with the control of quality of collected data, it was very important to provide the system with a certain degree of flexibility to allow the management of new possible tests, drugs and/or parameters. Emosys provides tools that permit simple database updating without the need to modify the program code. Another functionality by which Emosys supports the user is the possibility to associate a group of specific blood tests to a related disease. This is very useful when routine monitoring of blood tests for a particular disease is prescribed. In fact, it is not necessary to re-select the tests because the system recalls them automatically. The established association can be modified and updated at any moment by medical people. Taking into account that patients afferent to the HU of the University of Pisa are about one thousand/year and each patient can access the Hospital many times, it has been evident that particular care was needed to design the
system to manage such a mass of patients and to collect data in a correct and efficient way, in order to realize a well structured database. To achieve these goals the first step has been the modeling of the HU Department. In fig.1, links and priorities among the different parts of the clinical procedure are shown.

a) Laboratory’s diagnostic tests
b) Image-based studies
c) Specific medical examinations

The (a) group includes tests like CBC blood test, leukocyte differential count, electrophoresis and many others. It is possible to insert customized groups of tests without any restriction. The interface provides all the necessary tools to create, modify and delete tests at any moment.

The (b) group comprises studies like radiography, TAC, RMN, echography and so on. Images are recorded specifying which part of the human body has been examined and associated with the medical report.

The (c) group is made of medical examination like cardiological, urological, dermatological examination and so on, made by specialists. The system allows the insertion of new types of examinations into the system.

Emosys controls if the numerical values of the tests have a normal range or not and signals the eventual abnormality by changing the background color of the test. It is also possible to plot numeric type tests to show graphically the evolution of the parameter values, over time.

The follow-up of the patients is achieved by building their clinic history, at each examination. To add a new examination the doctor has to choose the current state of disease from an updatable list, the current Karnofsky index (the system displays also the corresponding Ecog index) and he has to insert the examination’s medical report. The system automatically associates the current therapy (if any) with the current examination. It is also possible to indicate the date of the next examination, possible support therapies and other notes. At the end, the doctor must sign the examination simply choosing his name from a list and the system can print a report for the patient.

The subsequent step is the diagnosis of the disease and its association with the patient. Emosys divides the diagnoses into primary and secondary. Primary diagnoses are associated with main diseases that in our case are, for example, Hodgkin Lymphoma, non-Hodgkin Lymphoma, etc., i.e. hematological diseases. Secondary diagnoses are associated with less important diseases (from a hematological point of view) like diabetes, osteoporosis and so on. When inserting a diagnosis it is possible to choose the histological definition of the disease, the current staging and, if available, notes on the beginning modalities of the disease, like initials signs and symptoms. Once the diagnoses has been established the medical staffs assigns a specific therapy to the patient. The management of the therapies is a very sophisticated task. Emosys provides the necessary tools to create a new chemotherapy, or a new non-chemotherapy, directly through the graphical interface of the system. A new chemotherapy is created in three steps. In the first step, the name of therapy, the recycle time, the number of administrations, the type of therapy (monoclonal or not) and other minor parameters are inserted. In the second
step, the administrations are named, the number of drugs for each administration is specified and the day distance between two administrations is decided. In the last step, for each administration, the drugs, the nominal dose for each drug, the unit measure, the type of the dose (pro-kilo, fixed, per m²), the days of administration, the method of administration and other minor parameters are specified. After that, the therapy is ready to be assigned to a patient. The system denies the assignment of more than one chemotherapy to the same patient. In other words, if there is an active chemotherapy for a patient, we cannot assign him to a new one. Emosys manages automatically all the steps involving the cycle of life of a therapy and allows to the doctors to search for therapies assigned to a patient according to several criteria. Then the system displays the found therapies in a graphical format so that it is easy to have a global vision of the ‘therapeutic’ history of the patient. Moreover Emosys lets the doctor manage many minors, but essential, parameters like drugs, samples used in laboratory, patient’s professions, hospitals involved in collecting the data, staging of disease, molecular biology tests, cytogenesis tests and others.

4. Data Collection

As it can be seen, the amount of data that must be managed is very high and these require an optimized implementation of the underlying database, otherwise the performance of the system would not be acceptable. The collection of all the clinical data must satisfy some very important requirements to be really useful. All the data have to be collected according to agreed protocols, in order to standardize the structure of the information and to assure their integrity. It must be possible to reconstruct the history of the patients and to register the outcome obtained. These requirements imply to build up structured databases, so that information retrieval algorithms, and new methods of data mining to discover new knowledge, can be applied. To achieve these objectives Emosys has been designed as a Web Application, accessible by any Web Browser, e.g. like Internet Explorer. Only authenticated persons can enter the system, by individual profiles. Privacy is granted by encrypting data over the network by the SSL protocol. The server manages web requests/responses and data encrypting through Tomcat Web Server and provides secure database transactions with Oracle 9i DBMS. The system supports Unicode UTF-8 for extended char set. In this way, it is possible to use special symbols when required, graphics chars and so on. Emosys offers a nice user-friendly graphical interface that facilitates the insertion of valid data. The types of data that can be collected may regard:

- Patient’s personal data
- Tests
- Examinations
- Diagnosis
- Diseases

- Therapies
- Clinical parameters
- Outcome results

4.1 Designing of the database

The design of the database has been carried out through three steps:

a) Meeting with the medical staff
b) Modeling of the Hematological Unit
c) Creation of the database schema

The first two steps (a) (b) follow from the modeling of the application domain. Then, as a result of this process, a functional UML scheme has been obtained. Based on UML diagrams we have created the entity-relation diagram that describes in detail the schema of the database. The schema contains the database’s tables, the relationships among the tables and all the necessary integrity constraints. We have created the Emosys schema using Oracle 9i DBMS. The database contains about fifty tables. The schema also contains the entire set of constraints assuring data integrity. For example, it is not possible to save in the Hospitals table two hospitals with the same name (unique constraints) or it is not allowed to delete a disease that it is associated to a certain diagnosis (relationship constraints). To manage concurrent accesses to the database tables we have used transactions, so if a command of a transaction fails the entire transaction is aborted and the changes are rolled back.

As we said before, the database is located in the central server of the Reference Centre. The clients send their data to the server through the Internet. To grant patient’s privacy the client encrypts any data before sending it over the Internet and the server decrypts the data upon receiving it from the network. We have adopted the 128 bit SSL protocol to manage data encryption. This protocol assures that no one who intercepts packed data can read the information transported.

The client performs several checks on the data being inserted. These local checks reduce the amount of computation of the main server, enhancing the performance of the system, and grant the conformity of collected data to the agreed protocols.

5. Patient retrieval and Test visualization

Emosys provides powerful tools to search for patients in the database. In particular the system implements a twofold level of search for patients: a simple level by generic search through the patient’s code, name, date of birth, fiscal code and so on, or a more complex allowing complex searching conditions on the following variables: hospital, age range, patient’s status (live or death), blood group, disease, stadium of disease, chemotherapy, status
of chemotherapy (active, terminated, stopped), hematological tests, image-based studies, and their combinations to express complex queries.

Another important feature of Emosys is the possibility to display simultaneously all the required tests and studies executed on a patient. This aids the user to compare different types of tests without the need to recall values or outcomes and thus decreasing the possibility of errors. For each category of tests or studies the system opens an autonomous window that can be managed independently of others. Respect to hematological tests, Emosys shows in different colors test that have values out of normal range, distinguishing values greater than the admissible upper limit (red) from values less than the admissible lower limit of the range (yellow). Normal range tests are black. The range is also displayed as the mouse pointer passes over the test field.

6. User management

One of the more important system services regards user management. Emosys supports a dynamic management of users in the sense that we can add or remove users in any moment, if we have the right privileges (Administrator). To add a new user in the system we have to specify a login/password pair, the name, the role and the belonging hospital. The role determines what actions the user can do and which information is allowed to view. There are three different roles:

- Administrator
- Doctor
- Assistant

The Administrator has full privileges and therefore can do any desired operations. The Doctor can do all the operations related to the management of the patient clinic folder, but he cannot manage users or alter very important system data. The Assistant can do only a subset of the operation allowed to a Doctor; in particular, he cannot alter parameters neither delete certain types of data, like an existing chemotherapy scheme.

7. Structure of the Web Application

Emosys is a Web Application running on Tomcat Web Server and interfacing with the Oracle 9i DBMS for database transactions. Typical client requests consist in database queries. Emosys executes the query, builds the appropriate response web page and sends it back to the client. The core of the system is the main servlet named EmosysManager, which receives the client requests, accesses the database through the JDBC Oracle Thin Driver, processes the data and then forwards the responses to the appropriate web resources that in our case are JSP (Java Server Pages) pages. The function of a JSP page is to build a customized HTML page depending on the client request. JSP pages are executed on the server side while the HTML they produce is executed by the client web browser.

Special efforts have been made to grant system stability in case of database access errors. A simple, but very efficient, optimization has been the use of a hash table to access the requested servlet methods instead of using a big cascade of conditional statements. In this way servlet methods are called in O(1) time against O(n) time in case of traditional if-then-else cascade method. This enhances the system performance raising the number of requests processed by the unit of time.

Emosys also manages user sessions. When a user logs in the system, the servlet creates a session for the newly connected user. The session stores several type of fixed information like connection information, and serves as a buffer to store the data coming from the executed SQL queries. This data is then used in the appropriate JSP page to build the client response. From a logic point of view, Emosys can be structured in several logic modules each performing a certain task:

- Personal patient management
- General tests management
- Diagnoses management
- Therapies management
- Examinations management
- Data Mining

For each logic module, Emosys provides a nice user-friendly graphical interface to perform the necessary actions. Extensive use of HTML and DHTML has been used to build up the interfaces as clear as possible.

8. Data Mining

One of the most important Emosys modules is the Data Mining module. As we stated before, the main objective of the Emosys project is to analyze the collected data in order to extract potential new knowledge or confirm current observations and experimental data.

In general, the main approaches to Knowledge Discovery can be divided into two categories: statistical methods, based on pure statistical analysis and data mining [1] methods that use specific algorithms to search for new knowledge in a database.

The Data Mining module of the system can be properly viewed as a dynamic module, in the sense that it is possible to apply several data mining techniques either based on classical statistics or on existing tools or customized tools. In particular, in preliminary tests using the Data Mining module based on an unsupervised neural approach (SOM) [2], interesting results has been obtained. The principal aim was to find new potential relationships among data starting from unknown ones and for this reason the unsupervised approach based on SOM neural networks to cluster data has been chosen. A Self-Organizing Map (SOM) [3] defines a 'non-linear projection' of the probability density function of the high-dimensional input data onto the two-dimensional display.
The network places a number of reference vectors into an input data space to approximate its data set in an ordered fashion. To implement the module the SOM-PAK tool, developed by the SOM Programming Team of the Helsinki University of Technology, Laboratory of Computer and Information Science, has been used. The data set at disposal it has been extracted from an old database of the Haematological Unit of the Santa Chiara hospital of Pisa and it has been chosen to develop and test the module. The data used are relative to about three thousand random patients. According to the KDD process, the first step has been data cleaning. In this step, inconsistent data has been discarded and valid data has been pre-processed and converted in the right network format. In particular, a vector quantization algorithm has been applied to the clinical data to obtain consistent data.

Then the system has been tested to find known results, by the medical point of view, to evaluate the goodness of the results. The next step has been the selection of a subset of attributes as characteristics of the training set for the neural network. The components that have been selected are patient’s city and patient’s profession to search for relationships among these data and different types of haematological pathologies. Then, the network has been trained and the resulting map has been graphically analysed with the Viscovery SOM Viewer (fig.2). The data set has resulted to be grouped in three distinct groups (clusters). A further study of the nature of the elements belonging to the same cluster has evidenced that one of these groups was in large part composed by farmers. This kind of classification has addressed studies towards an investigation to find common features in farmers. It has been found that farmers have a greater incidence in two particular diseases: Non Hodgkin Lymphoma and Chronic Myeloid Leukaemia. In fact, in the set of clinical data (3076 patients), the incidence of Non Hodgkin Lymphoma was 30.0% and for Chronic Myeloid Leukaemia was 4.6% while in the farmers subset (44 patients) was 42.2% for the first pathology and 28.9% for the second one. Medical literature [5] has confirmed that there is a potential relationship among the use of pesticides or other chemical products and the Non Hodgkin Lymphoma. In the second case, there is not a literature confirmation, but only preliminary results. Then this result can be used to address medical research to find new confirmation. The developed module is only a preliminary Data Mining application. In the near future, it will be applied to the new database under construction at the Haematological Unit of the Santa Chiara hospital of Pisa by using the Emosys system.

![fig 2. The map shows the clustered data set after training the neural network.](image)

**Conclusions**

Emosys is a flexible, platform independent, updatable, secure modular system. Emosys allows very sophisticated procedures in a very friendly approach and is now under validation at the clinical Sites and in a near future it will be possible also to show the first results extracted from the new database.

**References**


