Management of medical information through Adaptive Searching Mechanisms

E. de la Vega, S. Chessa, E. Gaeta, M.T. Arredondo
Life Supporting Technologies
ETSI Telecomunicación UPM, Spain
evega@lst.tfo.upm.es, schessa@lst.tfo.upm.es

Abstract

Medical knowledge is inherently complex and uncertain. Medical experts may provide different interpretations for symptoms since all of them also depend on a given context and most of them are established by statistical utilization. It is necessary a whole knowledge baggage in order to understand and take care of patients with cardiovascular diseases adequately. For this reason, to reach this goal, and due to the fast growth of the Web and its number of users, it is essential to develop user adapted systems based in Web environments. In this context, a platform has been conceived aiming at being a valuable instrument for cardiologic medical information retrieval from heterogeneous, distributed medical databases that mediates medical decision of critical health conditions.

This paper presents a new adaptive searching mechanism developed using innovative technologies in order to obtain, use and manipulate medical information.

The adaptation features that are supported in this platform generate a personalized searching process for the users depending on the information stored in their personal profiles. The technology chosen to store this profile information is XML owing to its portability and because is a structured language; to manage all the information the best solution is a XML native database. In this way, it is possible to extract the information directly in structured XML and modify directly the files stored in the database without any transformation. The main technique to set the platform profiling system consists on the generation of the most approximated suitable roles segmentations in order to identify users with the profile that has been generated of them. Over this segmentation, different layers of user's characterization are defined to refine the user personal definition. The most important aspect of this adaptive system is what is called dynamic profiling which is achieved through two different ways. Using advanced methods of capturing the user's activity and discovering which information items attract the user's attention by means of attentive agents. The structure selected to store information in these dynamic profiles is a Semantic Net built through concepts/terms that belong to patterns obtained with the methods mentioned before.

The searching process that takes place within the platform has the purpose of improving the queries introduced by the users of the platform, creating an advanced query that will be the base for the personalization of the results presentation. This feature aims at the management of the retrieved results in response to the previous query delivery. The results will be ranked and divided calculating the relevance to the user. In this way it is possible the integration between the information retrieval system and a graphic user interface, that is developed using Java Applet technology due to its portability, usage facility and possibility to execute searching application locally. To increase the system accessibility and to facilitate the integration, the functionalities are being implemented using Web Services technology, reaching a total portability.

1. Introduction

Medical knowledge is inherently complex and uncertain. Medical documentation can provide authorized persons with all relevant information about one or several patients and their treatment, at the right time, at the right place and in the right form. The primary goal of medical information is to support the quality management of a health care institution, especially by providing the most appropriate information. But medical experts may offer different interpretations for several symptoms since they depend on a given context and most of them are established by statistical utilization. It is necessary a global

knowledge experience in order to understand and take care of patients adequately.

It is vital to provide medical professionals with acute data on time, thus a tool that uses search mechanisms to discriminate between different treatment options and takes into consideration other possibilities is needed.

Nowadays there are excellent search engines able to find almost all kind of information, including medical one, inside an internal database or the Internet environment. Nevertheless, the ideal search engine still does not exist. One of the limitations of current search systems is that all users and queries are treated in a similar way, using a single strategy to retrieve information. However, medical information retrieval systems should respond differently in different situations to reflect the concrete needs of the users. Additionally, the fast growth of the Web size and its number of users makes essential the development of user adapted systems based in Web environments. Finally, users' behaviour should be the centre of attention of a search engine quality studies. There have been some studies on this topic, mainly focused on general users, but research has to be extended to certain user groups (i.e. how information professionals or members of a certain occupational group use or would like to use search engines).

In this context, this paper presents a new way to manage medical information using innovative search mechanisms. An Adaptive Searching Mechanism that takes user preferences into account has been developed. The system has been conceived to be a valuable instrument for medical information retrieval from heterogeneous, distributed medical databases and to help health professionals to take decisions. It has been developed focused on the cardiovascular medical field, due to its great social importance and economical aspects. Nevertheless, the tool could be easily adapted to support any other health area.

The solution proposed to encourage health information inside the medical environment is part of NOESIS, an Integrated Project partially funded by the EU. Its main objective is to develop an intelligent environment that enables ubiquitous management of citizens' health status and assists health professionals to answer their everyday clinical and scientific questions integrating them into clinical practice of advances in evidence based medicine.

2. Methods

The developed adaptive searching mechanism can be divided in two parts: a Search Module, with a Search User Interface based in Java Applets, and a Management Core, based on the interaction between Web Services and internal algorithms. The second module has been designed to manage the user profiles.

The Search Module is in charge of handling all the process since a user introduces a query into the system until the results and any other extra information are shown. The queries introduced by the users are improved to create an advanced query that will be the base for the personalization of the results presentation. Final results are ranked and divided calculating the relevance to a particular user. Therefore, it includes the necessary algorithms to capture a user query, to manage halfway procedures, data processing included, and finally to display results.

The aim of the adaptive searching mechanism is to provide health professionals with trustable medical information. This makes necessary the existence of different modules for storing information, indexing the diverse kind of documents information and extracting terms from queries written on common language. Additionally, the search system has to exchange information with several modules inside a global management system, the NOESIS platform, modules that have been developed using different technologies. The adaptive searching mechanism has been designed as an application that runs in the user device; thus, it implies the need of having a multiplatform configuration. Also, the search engine has to present graphical information to the user through a graphical interface. Hence, Java Applets have been selected as the appropriate technology for designing this interface (i.e. it allows executing searches without program's installation). The search module then has been implemented as a Java applet that connects all the required modules of the system through different Web Services clients, process the information and presents, through a GUI embedded into an HTML page, the results of the internal process. Finally a digital sign has been included in the system to provide security and confidence.

The user profiling module has been designed with the aim to automate the detection of the user's behaviour. It includes information such us demographics, language and search preferences, and personal interest in medical area stored during the sessions of each user in the system. The design of the module has been divided in two parts: the static and the dynamic one.

The static part of the profiling system contains data directly introduced by the user in the registration session on the system. This information is used by the search module to personalize the search interface and improve search results and also used for the interface

adaptation based on the user's preferences. The interface has been designed to provide a complete accessibility and to be user friendly. The store procedure of the segmentation of the suitable roles is used to identify the users. Over this segmentation, different layers of user's characterization are presented to refine the user personal definition.

The dynamic part of the profiling system is based on the information captured by advanced user activity capturing methods, attentive agents, and by user mining methods. The aim of these methods is discover the most relevant information on which the users have paid their attention visiting web pages. With this data, usage patterns are created and sent to the profiling system where they are stored dynamically and automatically with minimal user input. The structure chosen to store this data is like semantic net (**Figure 1**) and the language XML. The XML file is generated, if the user does not have the dynamic profile, or dynamically updated whenever a new pattern arrives; each pattern consist in a group of url and word that are stored as nodes in the dynamic net. These nodes are linked to others nodes that belong to same patterns. When a new pattern is considered useful for the system it will be aggregated to the profile. Whenever a pattern is outdated, it will be removed from the system (**Figure 2**).

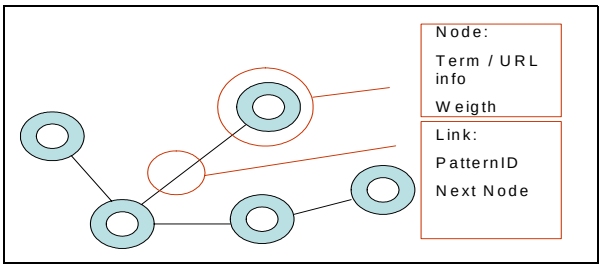

Figure 1. Elements of the dynamic networks

The aim of this semantic net is to provide the search module with a personalized expansion of the query made by the users (**Figure 3**). For achieving this goal an expand queries algorithm has been studied and implemented as a new method into the search module for the personalization of the searches.

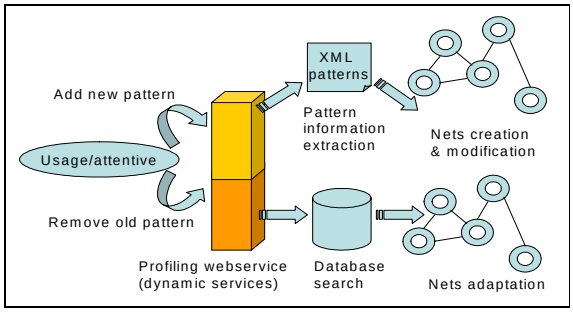

Figure 2. Queries personalization

Once the search module has previously extracted the relevant words from the originally query through a natural language process (NLP) algorithm, the first action was to decide if the order of the words queried was relevant or not. It was decided that the order is not significant. Thus, if a user searches three words (i.e. A, B, C) it will be the same expanding the query in any order. Finally, the algorithm implemented for expanding the user's queries (Figure 3) is mainly based on the following steps:

- First the algorithm calculate all the possible combinations of the query sent from the search module to the profiling system;
- Calculate the minimum route (minimize the jumps) that include all the query elements for each combination and then choose the minimum one;
- If there is only one minimum route, the algorithm chooses it for the expansion of the query else, if two or more routes have the same number of jumps the select is token through the results of two algorithms: page ranking (1) and hit rate;
- The hit rate algorithm estimates the hit rate through all the nodes that belong to a route. The hit rate of each node is a weight based on the times that a user searches the element. Hit Rate medium:

$$\overline{H}_r = \frac{\sum_{i=1}^{n} h_{r_i}}{n}$$

n= # nodes chosen in the route

hit rate i node normalized: $h_{r_i} = \frac{N_i}{N_{tot}}$,

$$N_{tot} = \sum_{i=1}^{m} N_i$$ m= # nodes in the net

The hit rate of each node is normalized with the total of the hits on the net due to avoid in the future overload of the values.

- In the same way the page ranking (1) medium is calculated:

$$\bar{P}_r = \frac{\sum_{1}^{n} p_{r_i}}{n}$$

Finally for selecting the route this formula is applied:

$$W_{global} = (1-\alpha)\bar{P}_r + \alpha\bar{H}_r$$

α = 0.8 is a parameter chose to give a high weight to the hit rate valour due to manage the aim of the system to expand and personalize the query.

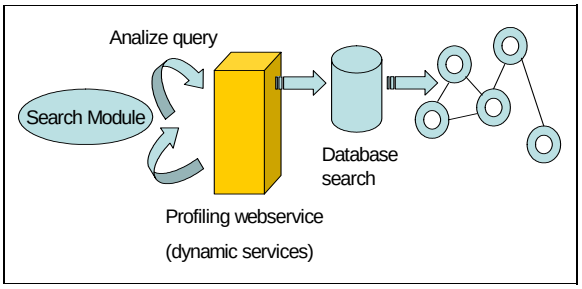

Figure 3. Dynamic profile query expansion

Different technologies have been considered and tested for being used in the development of the profiling system. Finally, the following ones have been selected:

- XML: for its portability.
- XML native database eXist: for its facilities to manage XML files.
- Web Services: for both their accessibility and integration facility.
- JAVA: Java is a portable programming language ideal for the development of network applications.
- XML parser: JDom and SAX.

3. Results

The general structure chosen for the Profiling Module is like a XML multi-platform access information server. It has a multi-access structure implemented with XML to be managed through the Internet.

The general configuration of the system consists of three modules:

- A Storage Module: responsible for the storage of all the information related to a user.
- A Management Module: responsible for the management of the information stored in the profiling system. It basically manages the profiles upgrading, the reports analysis and the access to specific user information.
- A Communication Module: responsible for the transmission and reception of all the information stored in the system.

The main functionality expected by the user when using this adaptive searching mechanism is the visual presentation of the results generated in response of a query inserted into the system. It is clear that is not enough with the simple generation of a list of links that the user will be able to click. To show all the possibilities promoted by a system based in the user profiles and in their experience, the simple results should be completed with a set of information elements that have to complement the basic information (Figure 4).

Besides these additional components, the general interface has been designed to adapt the information exposed to the singular user that is using the tool (i.e. all the labels´ language will be changed depending of the user mother tongue, and some additional tools will only be used by authorized professionals). These characteristics depend on some initial values that are established by the general platform where Adaptive Search is running. The communication is set through a main Search JSP page.

Something remarkable inside this innovative search mechanism is that images and text results are handled in the same multimedia mode. It can be traduced in the idea of medical users asking system for thoughts more than plain text solutions.

To understand the continuous process performed inside this adaptive search device, the search mechanism could be divided in three areas:

Query Improvement. When a user inserts a query in the system, successive processes are activated. Once the query has been converted into a useful terms succession, made by a Natural Language Processing Module (NLP), the information about user interests and preferences will be collected and added to the system. These last processes tally with the static improvement. Nevertheless, the last query modification has connection with dynamic area inside profile management: new terms will be added regarding to a continuous refresh of interesting data.

Information Ranking. There are some procedures that assist the system to classify the results depending on their relevance. It is done by means of User Profile mechanisms and indexed information attached to all obtained result.

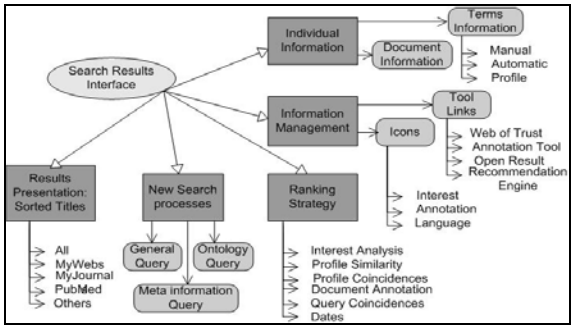

Figure 4. Functionality schema provided by the user search interface

Results' presentation. User will not visualize just a list of results, but also some additional features from other external resources available from this search interface (Figure 5). The information displayed on the screen and the set of functionalities offered to each user depend on the user type and his access rights. Main functionalities are:

- Five list's types depending on the result origin and user preferences. Text documents, graphical resources or combination between them are possible.
- Possibility of implementing new search processes beginning from last results.
- Ranking option can be update to have a better perspective depending on the user interest.
- Several buttons to communicate with other tools to annotate the information, to obtain recommended links and to mark a specific document or author.
- Visualization of some relevant metadata of each result, like title, author, language, publication's date, etc. Besides this there are some icons that facilitate a fast result comprehension.

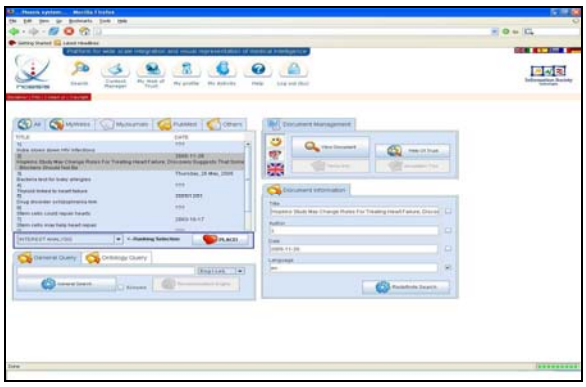
Figure 5. Functionalities over Search Interface

4. Discussion and conclusions

The clinical validation of the system has just started and there are not results yet. However, the system is having a good impact during the presentation to the users and it seems that the interfaces are user friendly and easy to use.

The integration of the functionalities has been very easy and without serious problems thanks to the web services technology used for the implementation. It has been also shown the efficiency in the tasks related to the profiling and searching processes described in this article. The personalization of the search mechanism, done by filtering the information retrieved using the personal user information, allows returning to user only the needed information avoiding time losses. In this way the system eliminate all the unnecessary information associated to a searching process and to adapt the information searched for a single user.

5. Acknowledgements

We would like to thank the NOESIS Project Consortium their valuable contributions for the realization of this work. This project is partially funded by the European Commission.

6. References

[1] Culver JD, Gerr F, Frumkin H. Medical information on the internet: a study of an electronic bulletin board. J Gen Intern Med 1997; 12: 466-470[Medline].

[2] Gunther Eysenbach, resident, Thomas L Diepgen, consultant in dermatology. *Towards quality management of medical information on the internet: evaluation, labeling, and filtering of information*. Editorial by Coiera. Unit for Medical Informatics, Epidemiology, and Public Health, Department of Dermatology, University Hospital Erlangen, Hartmannstrasse 14, 91052 Erlangen, Germany

[3] Irene V. Blair and Geoffrey R. Urland, University of Colorado, Boulder, Colorado and Jennifer E. Ma, University of Kansas, Lawrence, Kansas. *Using Internet search engines to estimate word frequency*.

[4] Dr. Dirk Lewandowski. *Web searching, search engines and Information Retrieval*. Information Services and Use 18(2005)3

[5] F. Leiner, W. Gaus, R. Haux and P.Knaup-Gregori, *Medical Data Management. A Practical Guide*, Health Informatics Series, Springer.

[6] Modeling and User-Adapted Interaction, 6(2-3), 131-155, 1996.

[7] Epidemiology, and Public Health, Department of Dermatology, University Hospital Erlangen, Hartmannstrasse 14, 91052 Erlangen, Germany

[8] Ryan C. LaBrie. *The Impact of Alternative Search Mechanisms on the Effectiveness of Knowledge Retrieval.* Arizona State University, May 2004

[9] Hugo Liu, Henry Lieberman, Ted Selker. GOOSE: *A Goal-Oriented Search Engine With Commonsense*. MIT Media Laboratory 20 Ames St., E15-320G Cambridge, MA 02139, USA

[10] Hubertus Hohl, Heinz-Dieter B–cker, and Rul Gunzenh”user. "HYPADAPTER: An Adaptive Hypertext S

[11] Modeling and User-Adapted Interaction, 6(2-3), 131-155, 1996.

[12] SUN Microsystem: http://www.sun.com/java

[13] NOESIS project “Description of Work”, IST 507960. European Commission Framework Program