

METIORE: A Publications Reference for the Adaptive Hypermedia Community

David Bueno¹, Ricardo Conejo¹, Cristina Carmona¹, Amos A. David²

¹ Department of Languages and Computer Science, University of Málaga,
29071, Málaga, Spain.

bueno, conejo, cristina@lcc.uma.es

² LORIA, BP 239, 54506 Vandoeuvre, France
adavid@loria.fr

Abstract. The Web is one of the most powerful sources of information on any topic. However looking for scientific literature is a difficult task. Prior knowledge of link sites is necessary and if you are lucky they point to conferences proceedings available on-line. In fact the case the user is not able to make queries about the available documents and must check them one by one using general purpose search engines. In this paper we propose our system METIORE as a source of information for the Adaptive Hypermedia community. The idea is to put together all the publications on this research area and provide an adaptive tool to find papers or people working in the field. METIORE is a Personalized Information Retrieval system that keeps a user model based on objectives.

1 Introduction

The medium most used by the research community to find relevant work is the Web. The search engines such as Altavista or Google are improving their way of retrieving information. These engines index Web documents, but many of the scientific publications are in postscript or pdf formats which are not indexed by the engines. Also, because of the large quantity of Web pages they index, many of the results that can be obtained are not relevant. It is therefore necessary to provide information resources and associated tools where researchers can just find scientific contents without the usual noise introduced by typical search engines.

The idea of having a search engine exclusively for scientific information has already been developed. FermiVista [3] is an index of articles and prepublications available on the Web or FTP servers of Universities and Research Centers. It contains documents in the domain of mathematics, computer science and physics. Unfortunately this index has not been updated since 1998.

One of the best specific indexes is the CiteSeer [4]. It locates scientific articles on the web using search engines like AltaVista and allows for the extraction of information of the citations in publications. The main feature is the creation of an

Autonomous Citation Indexing [5]. Even though this system can be applied to any field of research, it is specific for Computer Science literature.

In this paper we present the application of our Retrieval System METIORE oriented to the implementation of an index/search engine for the Adaptive Hypermedia (AH) publications. In section 2 the generalities of METIORE will be presented. The adaptation to the Web to make the AH index is explained in section 3. Finally the conclusions and future work are presented in section 4.

2 METIORE

METIORE is the acronym of Multimedia coopErative InformatiOn Retrieval SystEm. It is a multipurpose IRS that can be applied to different kinds of database. It has been used in three different areas. For the first area, METIORE_STREEMS is an IRS for managing multimedia information on trees authorized for reforestation by the European Union (EU). The project was sponsored under the EU project LEONARDO. For the second area, METIORE_LORIA is used for managing the database of publications of the computer science laboratory research center, LORIA, Nancy France. Some evaluations of METIORE have been done using this database [1]. In the third area, METIORE_REVUES is used for the access and analysis of a collection of journals called "Relations Publiques Informations".

The ability of METIORE to manage different types of data is in part due to the use of XML to import the different databases. From these sources it generates files with inverted lists and cluster information to provide a fast access. The core of METIORE is independent from the data; it only needs to know the parameters that will be used for the personalization. So if it is working with publications it may use author, title, keywords, etc. But if it is working with data on trees, the parameters may be height, altitude, scientific name, etc.

METIORE has a powerful search interface that allows the user to make simple or complex queries and also perform global analysis of the database. It is also possible to explore the database without knowing anything about its content. This last possibility is very useful as a starting point in the process of information retrieval. The following are examples of the types of query that can be processed using METIORE on the database of publications:

- Documents about Script languages published after 1998
- List of all authors with their number of publications
- Authors who have written with X and how many times (this shows people working together)
- Names of authors and keywords that they use
- The yearly frequency of the use of the keyword 'expert systems' (this provides a means of observing the evolution of a research topic in terms its frequency of use in publications)

The most relevant feature of METIORE is the personalization of the response. Our approach on personalization is based on the concept of objective. For us the objective is the expression of the user's information needs formulated in natural language. The

objective is currently used to group the set of queries, concepts and decisions that the user makes on the system having his objective in mind. Our hypothesis is that grouping the user's interactions into objectives will help the user to find information in his history and help the system to build a specific model of the user from one or more sessions.

This concept is very important to us since we believe that every user has a minimum knowledge of his information needs before attempting to use an IRS. This knowledge can of course be improved with the use of the system and consequently the user's ability to express his information needs through the system's interaction. For us, the queries do not necessarily express all the information needs of the user but rather his approach towards solving the problem of his information needs.

For each user objective the system constructs a model. In order to give users a personalized response, after querying the system, they should evaluate some documents of the response. This provides the system with the user's preferences for this objective and thus help the user in future queries to have the most relevant solutions first. This is different from many retrieval systems that gives a list of solutions (for example 400 documents) that are sorted only according to the user's query. METIORE sorts the solutions using the query and the user model in order to select the best solutions that are presented at the top of the list.

Another interesting feature of METIORE is that the user has available his history of evaluated documents. This history is organized by objectives, for each of which the evaluated documents are sorted by their relevance to the user's objective. With the history the user can review his past solutions, modify his evaluations and also look for related documents.

3 METIORE ON THE WEB

To make METIORE more accessible, it has been necessary to make it available on the Web. The best option has been to keep the core of METIORE to maintain all its functionality and used as a server. The user will have a web interface that is connected to a *Java Servlet* which will send transactions to the METIORE daemon through a socket connection. A specific protocol has been created for the communication between the applications. The main programming language of METIORE is *incr Tcl*. Fig. 1 shows the architecture for METIORE on Web.

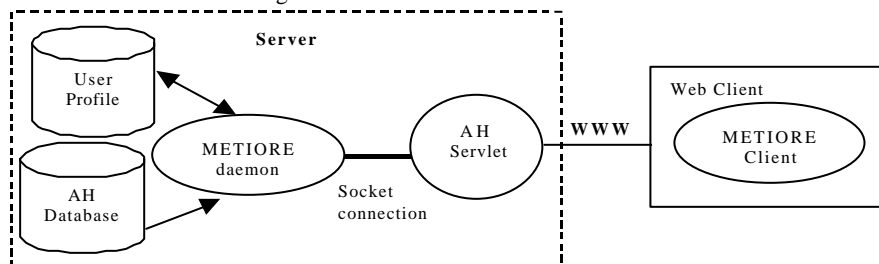


Fig. 1. The architecture to make METIORE accessible through Web

For Adaptive Hypermedia literature and in general for scientific publications on the Web we have designed a database that contains all the information needed to represent a publication (including its url). It also includes the information on authors and conferences/journals with links to their homepages. This makes METIORE_AH a source of data where papers, conferences, editorials and people can be easily found and contacted.

All the publications of the Adaptive Hypermedia Home Page [2] and related conferences are included in the database. We have realized that most of the papers don't have associated keywords. In order to obtain the keywords (more precisely the root of keywords) and have them as a source of information in queries, the full documents are treated in the following way: The paper is parsed using a stop words list and the porter algorithm [6]. Then the frequency of words is obtained and the most relevant are chosen to represent the article. A web page is also provided to let researchers include their publications in METIORE. When all these documents are inserted, it will be possible to use the characteristics of METIORE to obtain relevant publications for the users through the Web.

4 CONCLUSIONS AND FUTURE WORK

In this paper we have presented the application of METIORE for managing the Adaptive Hypermedia publications on the Web. We hope to integrate some other functions that are available in similar systems such as CiteSeer. One of the functions is the processing of the references to have some kind of citation index, or to automatically include papers to the database. However, METIORE offers other possibilities that can't be found in other applications, like the complex data analysis, the objectives of retrieval that can be maintained in different sessions, or the active history to review relevant publications. Others features such as having the list of the most relevant articles or authors according to the interactions of users with METIORE will be available soon. With the data obtained through the user's interactions, new evaluations of METIORE will be made. Our objective is to apply METIORE to other areas of research after evaluating this application.

5 REFERENCES

- [1] Bueno, D., & David, A. A. (2001). "METIORE: A Personalized Information Retrieval System ". User Modeling 2001
- [2] De Bra, P. (1999) Adaptive Hypertext and Hypermedia [Web Page]. <http://www.wis.win.tue.nl/ah/>.
- [3] Fermigier, S. (1996) FermiVista [Web Page]. <http://fermivista.math.jussieu.fr/>.
- [4] Lawrence, S. (1997) CiteSeer Research Index [Web Page]. <http://citeseer.nj.nec.com/>.
- [5] Lawrence, S., Lee Giles, C., & Bollacker, K. "Digital Libraries and Autonomous Citation Indexing". IEEE Computer, 32(6), 67-71
- [6] Porter, M.F., (1980). "An algorithm for suffix stripping", *Program*, **14**(3),130-137