Personalized intelligent information services within an online digital library for medicine: the BIBLIOMED system

Paolo Omero*, Nello Polesello†, Carlo Tasso*
*University of Udine
Email: omero@dimi.uniud.it
tasso@dimi.uniud.it
†infoFACTORY
Email: polesello@infofactory.it

Abstract—The services available in current state-of-the-art digital library systems may profit from new technologies capable of increasing the level of personalization and the accuracy in the retrieval. In a longer perspective towards the future, it is foreseeable that new services will be added to digital libraries, going beyond simple searching and retrieval, adding new levels of processing aimed at automatic information extraction, analysis, and organization. Going even further, the digital library can be seen as a part of a knowledge-based community, and this may call for new even more sophisticated tools. We present here the ongoing BiblioMed project which proposes a new open model for the organization of services offered by a digital library, aimed at providing more effective information access, in terms of accuracy, personalization level, timeliness, integration, completeness, manageability, and cognitive affordability, overcoming in such a way some of the current limitations. BiblioMed has been developed within the Medical Library of the University of Udine.

I. INTRODUCTION

Common state-of-the-art tools for accessing information in a digital library within a medical research and healthcare center include standard retrieval systems such as online catalogues and specialized search engines provided by several commercial service aggregators (Blackwell Publishing, Elsevier, Kluwer, CILEA, etc.): doctors, scientists, and other health professionals use these tools for accessing scientific and technical papers published in subscription-based online journals. Another common approach is the exploitation of public online services like PubMed or other medical (vertical) portals available on the Web. Most of such tools feature a key word based approach for searching (notoriously known as a very low precision mechanism) and a very low level of adaptive personalization (known as a key factor for improving quality of service) [Brusilowski & Tasso, 2004], [Tasso & Omero, 2002] . From a more general point of view, the major limitations which characterize nowadays services provided in a digital online library fall into two perspectives: user perspective and library manager perspective. Let us consider the first ones:

L.1. Lack of personalization in accessing the information [Renda, Straccia, 2005], [Brusilovsky & al., 2005]: users are commonly treated all in the same way, even though they are characterized by very different information needs and short/long term interests [Mizzaro & Tasso, 2002]. A personal profile could help in filtering out information, overcoming in such a way the severe problem of information overload, whose effects are particularly evident in the medical field.

L.2. Lack of a proactive autonomous information service capable of detecting among the huge amount of (new) available information only the information relevant for the single specific user, without the need of explicitly requesting for it or formulating precise search queries.

L.3. Lack of timeliness in being informed about new available information, unless the user has the time to undergo a time-consuming and cognitively-demanding process of manual (periodic) search.

L.4 Low coverage, with respect to the very many sources available in the medical field: hundreds of online sources are available on the Web, including directories, blogs and forums, devoted to different kinds of users, ranging from scientists, to practitioners, to patients. Even though these resources increase the information overload situation, more and more they are recognized as very important references, to be exploited together with more traditional ones.

L.5. Very low (if none) integration level among the available resources, which obliges the user to utilize very many different systems, often repeating the same search (with the same queries) in different online systems (with different user interfaces) [Rao & al., 1995].

L.6. Waste of information and no empowerment of the available resources, caused by all the points above: the user does not know about a useful paper or a useful information source or does not have the time nor the cognitive energy for carrying on manually all the needed activities. A typical effect is that some subscriptions to online journals are not sufficiently exploited.

The library manager perspective is crucial as-well, especially for an optimized use of financial resources and for

providing the conditions for high quality services. We can mention:

L.7. Lack of control over the use of subscription-based online journals. Library managers are often unable to have a clear vision (over time) of the journals that are actually accessed, of how much they are accessed, etc.

L.8. Lack of information about the real interest of the users for online (subscription-based) resources not yet available in the library, and whose subscription can not therefore be easily recognized and evaluated.

L.9. No integration between search services and typical interlibrary services (such as interlibrary loans or interlibrary document delivery), which usually requires manual intervention of the library personnel.

The limitations described above are the result of an informal analysis carried out at the University of Udine and may be easily generalized outside the medical environment, to digital libraries in other disciplines.

In this paper we propose an innovative integrated model of library services, called BiblioMed, aimed at overcoming the above limitations, providing in such a way more effective information access, in terms of accuracy, personalization level, timeliness, integration, completeness, manageability, and cognitive affordability. Our proposal stems from the innovative tools and technologies developed over the years in the Artificial Intelligence Laboratory of the University of Udine for the semantic analysis and the personalized selection of (textual) contents. The BiblioMed model has been developed within an ongoing project carried out by the Faculty of Medicine of the University of Udine and the Udine Hospital.

II. OVERALL ORGANIZATION OF THE BIBLIOMED MODEL

The resources and systems connected to BiblioMed as basic information providers are:

- medical data banks, such as PubMed for accessing Medline;
- online medical journals, which may or may not require subscription;
- service aggregators, providing access to specific sets of resources, such as online journals and data banks;
- online bookstores available on the Web (such as Amazon, B&N, etc.);
- library catalogues (both internal and external to the university, located in other libraries);
- vertical Web portals, specialized in medical topics;
- medical online forums, blogs, and directories.

BiblioMed allows the continuous monitoring of all the available resources and the discovery (and integration) of new medical resources, taking into account personal needs and interests of single individual users. These capabilities are obtained by means of a Web content-based monitoring and filtering module called *ifMONITOR (information filtering Web Monitor)*, based on the ifT system [Asnicar et al.,

1997] featuring intelligent retrieval and filtering techniques. ifMONITOR feature various (adaptive and adaptable) personalization techniques, exploiting semantic networks and shallow (lexical-level) natural language processing for analyzing and representing the meaning of textual content and the user's information needs *(semantic profiles)*. A specific module is devoted to *link resolving*, exploited in order to verify whether a document (a paper) found by the system or by the user through a search service is accessible in the full text version, and in such a case, to provide a transparent access. If the full version is not available, services like interlibrary *document delivery* are possibly activated. All the different sources are integrated, and access is uniformly provided. The users can use a *Journal Search Engine* to find a specific subscribed journal providing the title, MeSH[1] terms or the editor's name. A *Meta Search Engine* allows users to make a query to all the selected online resources. Specific services are devoted to library managers and operators, which can monitor the usage of the available online resources, analyze the requests for service submitted by the users, and manage the interlibrary services. Fig. 1 shows the overall architecture of BiblioMed.
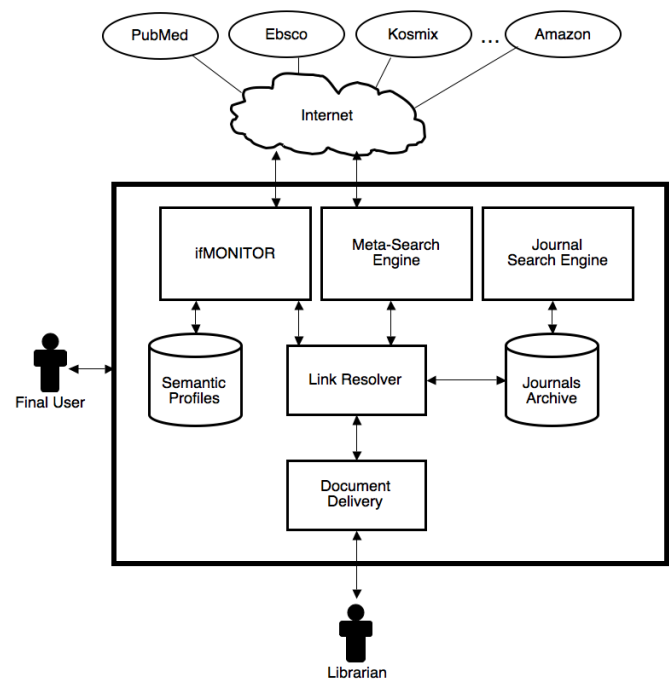

Fig. 1. Architecture of Bibliomed.

III. BIBLIOMED SERVICES

Beside the availability of basic library services (access to the catalogues, direct access to the subscribed journal/data banks/etc., and consultation), the major and more innovative services available in BiblioMed are the following:

---

[1]Medical Subject Headings

1) *Adaptive personalized service for monitoring and filtering any of the available resources, on the basis of a personal interest profile.* The profile, represented through a semantic (co-occurrence) network, is automatically built through sample fragments of texts, describing the topic of interest. Processing of the documents includes a lexical analysis constituted by (i) stop word elimination and (ii) lexical stemming. Documents representations are then matched against the user interest profile and the documents reaching a matching score below a given threshold are filtered out, the other documents being collected in a specific searchable archive. Thank to the amount of information used to represent the user need which is much higher than the amount usually present in a single key-word based query and, moreover, thank to the contextual (co-occurrence) information contained in the semantic network allowing to disambiguate the word senses, the negative effects of polysemic terms are highly decreased and the resulting precision is very high. Fig.2 shows a profile about "cataract multifocal intraocular Lens (IOLs)". The representation of the semantic network is a list of most relevant terms of the specific topic of interest and the most important relations of co-occurrence among terms. Fig. 3 shows the semantic network presented as a triangular matrix. Darker colors represent the strongest co-occurrence relations among terms.
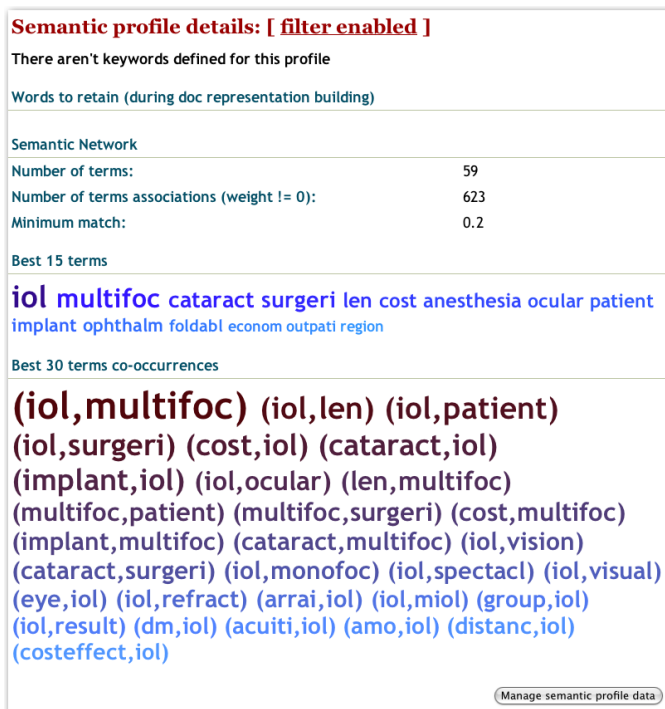

Fig. 2.   Representation of the semantic network as list of best terms and co-occurrence relations.

2) *Adaptable (customizable) SDI (Selective Dissemination of Information) service by means of a meta-search*
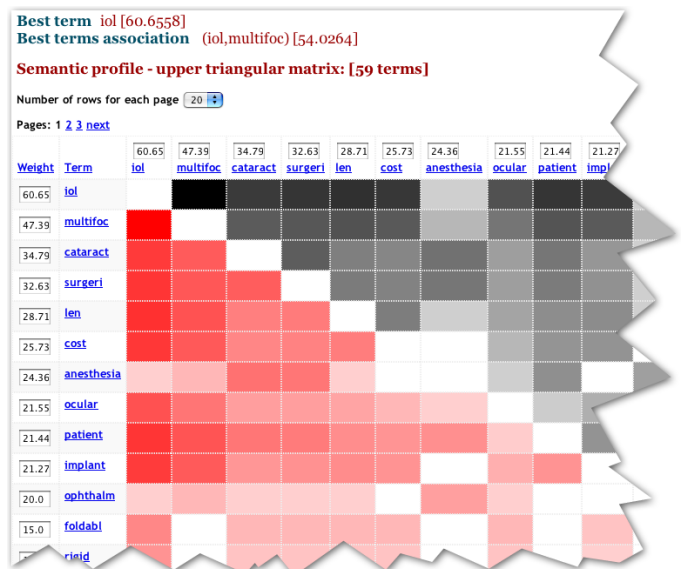

Fig. 3.   Representation of the semantic network as a triangular matrix.

engine, connected to several online resources, such as medical data banks (like PubMed), vertical portals, specialized search engines, etc. Each different user can select the resources to be searched and can provide the query to search; the meta-engine submits users requests to the selected sources, either once, or repeatedly (for example every day, or every week), according to the configuration proposed by the single user. The collected results are assembled and duplicates are removed. Since the meta-engine is based on key-word based queries (with possible low precision caused by polysemic query terms), the user is given the possibility to request the ranking of the retrieved documents by means of one of his personal (semantic-network based) interest profiles. Fig. 4 shows the meta-search engine and a list of results. Fig. 5 shows two results of the search. For each result the system shows the link to full text version (built by the Link resolver module) or the link to the Document Delivery service if the full text is not available.

3) *Selective alerting* [Buchanan, Hinze, 2005]. The user can request to be alerted (via e-mail or pop-ups) of any new document which is retrieved in one of the two services illustrated above.

4) *Journal search engine.* Given the huge number of online journals, a specific search engine is provided, which is able to work on the journal titles, MeSH terms and possibly on an extended description of their content. In this way the user can find out specific journals which he did not know (or remember) and which he may consider worthwhile to access.

5) *Personalized (user-customizable) utilities*, such as personal folders, links, local searching (on personal archives, i.e. those archives produced by the service mentioned in pt. 1. above), import/export functions, pdf

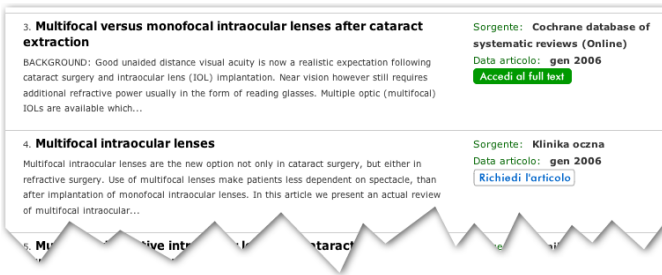Fig. 4. The user interface of the Meta-Search Engine.
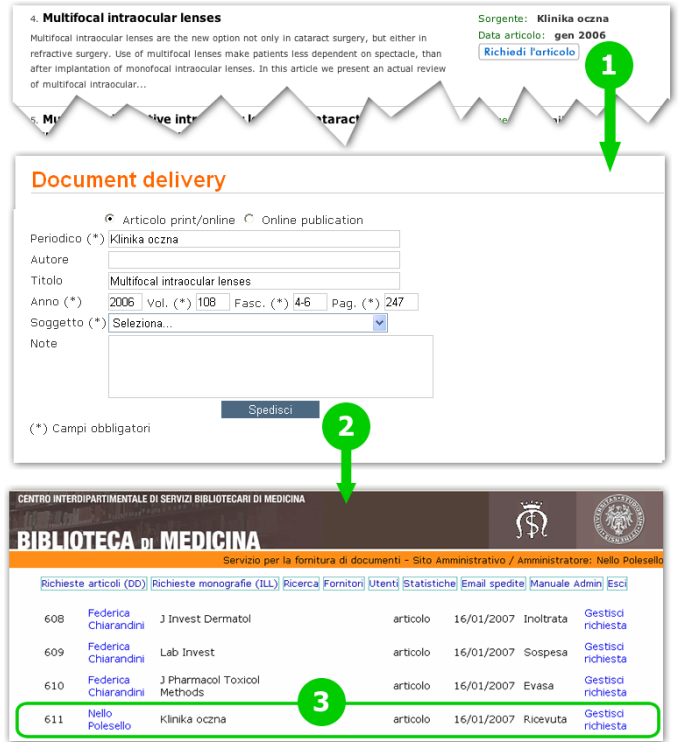

Fig. 5. Results from the Meta-Search Engine.

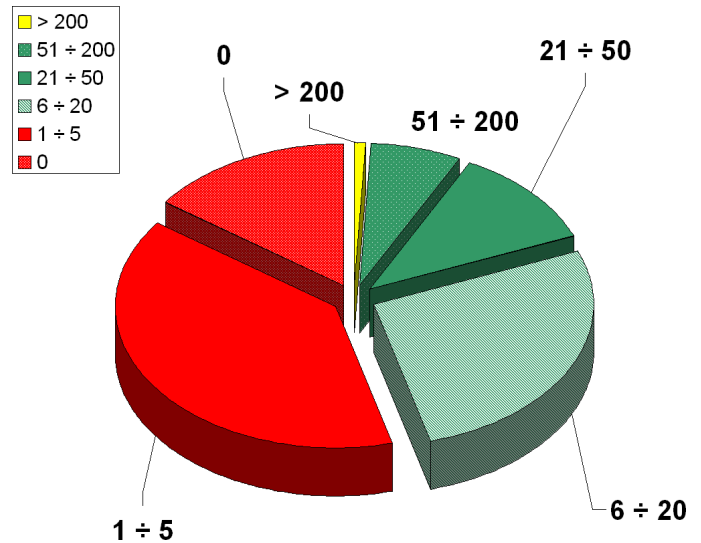Fig. 6. Integration with the Document Delivery Service.

Fig. 7. Number of journal accesses in the last six months.

conversion, personal annotation and evaluation, etc.

6) *Integrated interlibrary document delivery and interlibrary loan services*, which can automatically be activated whenever a specific paper is not available (in full version) in the local library: in this case it can be requested to another library. Fig. 6 shows (1) the link between a result not available in the local library and (2) the form to request the paper by Document Delivery Service. The form of DD service is filled automatically by the system and connected to (3) the interface used by the librarian.

7) *Monitoring of accesses to the online journals*, allowing the library managers to get a measurement (and its evolution over time) of the usage by the users of the different journals, providing useful information for strategic management and planning.Fig. 7 shows a statistical representation of journal accesses. We can see that 1024 journals (that is 54% of the total) are used less than 5 times in the last six month.

IV. PROJECT STATUS

The first version of the BiblioMed System has been released in June 2006 and is currently used by over 1000 users at the University of Udine and at the Udine Hospital. The analysis of the actual usage of the system is ongoing and it exploits a set of indicators which include: the number of personalized search profiles created by each users; the number of retrieved papers; the precision of the filtering system; the type and number of the accessed resources and

services (including most accessed ones and seldom accessed ones), the total (estimated) cost of the less (or never) accessed subscribed online journals, etc. We are also identifying the most active users of the system with the aim of carrying out a series of interviews for a joint assessment of the effectiveness of the system, its usability and for obtaining a continous feedback on the needed improvements and requirements. The first version of BiblioMed features the basic services, whereas the completion is planned for the next spring. At that time a full evaluation will be carried out.

V. Expected Results

Among the main expected benefits deriving from the use of the BiblioMed system, we highlight the following:

1) *Better exploitation of available information resources, time saving and higher productivity.* The ifMONITOR module in the BiblioMed system, frees the users from a heavy activity of manual search of information and provides new and only relevant papers timeliness. Without this kind of automation the user could give up accessing the available resources for lack of time.

2) *Higher flexibility in accessing various online information resources in an integrated and uniform manner.* BiblioMed is not bound to a specific set of information resources and can be easily extended with new ones.

3) *Easier use of traditional services.* The high level of integration among various resources and the link resolver module allow the users to avoid the traditional manual search process, the verification of the availability of the full text version of the paper and the operational effort required to use the Inter Library Loan service. Moreover, uniform homogeneous access frees the user from the need of frequently learn new procedures for operation.

4) *More rational management of subscriptions.* The librarians can more easily retrieve a list of the least used journals and assess the opportunity to substitute them with other resources.

VI. Future Works and Conclusions

With BiblioMed we claim to have partially overcome the traditional limitations of accessing information in digital libraries. The project is ongoing, and a systematic evaluation of its impact is planned and it will be reported later on. In terms of future developments, we are currently focusing on the extension of the BiblioMed model to other disciplines (outside medicine) where a centralized system like Medline is not available. In this case it is often necessary to access various heterogeneous data banks and an integrated access (as the one provided by BiblioMed) could be even more appreciated. We plan to extend the BiblioMed model with two types of services beyond retrieval:

1) extending the variety of automated tasks by processing the retrieved information in order to perform information extraction tasks (e.g. automatic identification of the

relevant authors in a specific research area [Tasso et al., 2005];

2) moving towards knowledge management activities by adding advanced cooperative work and community tools in order to exploit, sharing and distribute knowledge and to promote and consolidate (possibly interdisciplinary) communities of practice [Tasso et al., 2004] .

VII. References

[Asnicar et al., 1997] Asnicar F.A., Di Fant M. and Tasso C. User Model-Based Information Filtering. In M. Lenzerini (Ed.) AI*IA 97: Advances in Artificial Intelligence - Proceeding of the 5th Congress of the Italian Association for Artificial Intelligence, Rome, I, September 17-19, 1997, Springer Verlag, Berlin, LNAI 1321, 1997, pp. 242-253.

[Brusilovsky & al., 2005] P. Brusilovsky, R. Farzan, J. Ahn, Comprehensive Personalized Information Access in an Educational Digital Library, JCDL'05, ACM Press, pp. 9-18.

[Brusilovski & Tasso 2004] Brusilovski P. and Tasso C. Preface to Special issue on User Modeling for Web information retrieval. User Modeling and User Adapted Interaction The Journal of Personalization Research 14(2-3), 2004, pp.147-157.

[Buchanan, Hinze, 2005] G. Buchanan, A. Hinze, A generic Alerting Service for Digital Libraries, JCDL'05, ACM Press, pp. 131-140.

[Mizzaro & Tasso, 2002] Mizzaro, S., Tasso, C., Ephemeral and Persistent Personalization in Adaptive Information Access to Scholarly Publications on the Web, Proceedings of AH2002 Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Malaga, Spain, May 2002, SpringerVerlag, Berlin, LNCS 2347, 2002, pp. 306-316.

[Rao & al., 1995] Ramana Rao, Jan O. Pedersen, Marti A. Hearst, Jock D. Mackinlay, Stuart K. Card, Larry Masinter, Per-Kristian Halvorsen, George G. Robertson, Rich interaction in the digital Library, Communications of the ACM, Aprile 1995, vol. 38, n.4, pp. 29-39.

[Renda, Straccia, 2005] M. E. Renda, U. Straccia, A personalized collaborative Digital Library environment: a model and an application, in Information Processing and Management, Vol. 41, Elsevier, 2005, pp. 5-21.

[Tasso & Omero 2002] Tasso, C. and Omero P., La personalizzazione dei Contenuti Web: E-Commerce, I-Access, E-Government, Franco Angeli, Milano, 2002.

[Tasso et al., 2004] Tasso, C., Rossi, P.G., Virgili, C., and Morandini, A., Content-based personalization for asynchronous communication tools: the ifForum system. In L. Aroyo and C. Tasso (Eds.) AH 2004: 3rd International

Conference on Adaptive Hypermedia and Adaptive Web-Based Systems - Workshop Proceedings, Technische Universiteit Eindhoven, Eindhoven, August 23rd, 2004, ISSN 0926-4515, pp. 324-330.

[Tasso et al.,2005] C. Tasso, P. Omero, A. Baruzzo, M. Cignini, A. Felice, A. Morandini, N. Polesello, M. Valotto, C. Virgili, Sistemi intelligenti per l'accesso all'informazione Web, il knowledge management e le comunita' virtuali: il progetto infoFACTORY, Atti XLIII Congresso AICA, Parte 1, volume 5-7 ottobre 2005, pp. 188-201.