Intelligent Document Retrieval from Heterogenous Cardiology Literature Sources: Use of an Automated Notification Robot Learning from his User

T Berger, H Kottkamp, J Ramming, CH Elsner

Heart Center Leipzig, Leipzig, Germany

Abstract

Information management becomes harder with the increasing flood of information. The user needs systems which take a preselection of the information provided and which report periodically to him on their own.

The described project retrieves its content from the Pubmed Knowledge-Base at the NCBI and diverse internal datasources of a in-house document-management solution and links to the corresponding (online) documents. Only for Medline a similar approach is performed by services like http://www.pubcrawler.ie/ and http://www.biomail.org.

The special feature of the here described work is the learning / grading function of the software, which allows the semi-automatic improvement of the user (search) profiles. Some of the basic components of this project are shared as Open-Source with partners.

1. Introduction

Information management becomes harder with the increasing flood of information.

Recents studies reveal, that knowledge in medicine doubles every 3 years [1]. The user needs systems which take a preselection of the information provided and which report periodically to him on their own. Similar approaches are already performed by http://www.pubcrawler.ie/ and http://www.biomail.org.

Additionally for such a system it seems feasible to use special user profiles to cross-check for similarities in searches and then use the profiles for cross-recommendations between the users. This way an economic benefit can be drawn form the generated profiles and knowledge can used more efficient. Similar approaches are already performed by tools like e. g. http://www.fooxx.com.

1.1. Objective of work

Objective of the Work was to design an Open-Source Bot Software to allow users not only a periodical email/browser update of certain PubMed Topics and internal resources classified in MeSH – the system should also be able to process recent user feedback and modify the user and system profiles depending on the feedback. This way a combination of the features of the described two system approaches should be set up. Figure 1 shows the basic workflow of the used system.

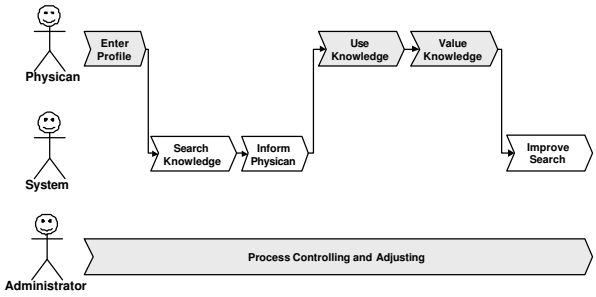

Figure 1. Basic Workflow of the Knowledge-Management System being used

2. Materials and methods

The described Project retrieves its content from the Pubmed Knowledge-Base at the NCBI and diverse internal datasources and links to the corresponding (online) documents. The system was designed as a simple web-interface, where users can login and configure their profile consisting of:

- Treated diseases and Applied procedures
- Their medical fields of work (from list of branches)
- Their interest areas (from a list of branches)
- Information sources (Journals / internal Documents)
- Keywords form MeSH

The internal resources were full-text indexed and additionally classified over the MeSH [4]. Additionally the selection of the resources is done over different keyword pattern associated with the user´s medical discipline(s) and the user´s department of work. Figure 2 shows the different associations between users, medical disciplines and their departments.

If new knowledge is found, the User is notified and asked for his input to grade the collected knowledge. Using this classification it is possible to pronounce the high-graded searches for a user and discard the low-graded ones.

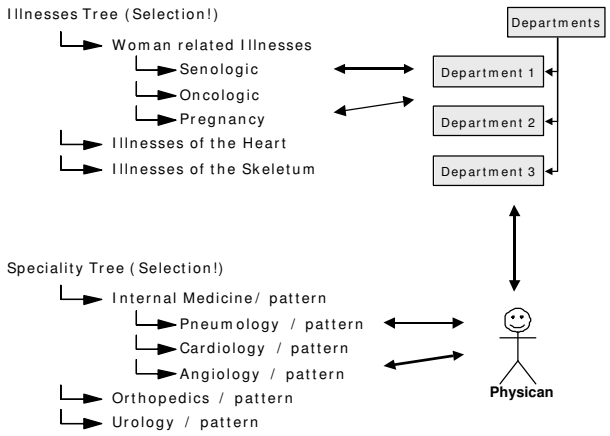

Figure 2. Pattern of Linking between Physician and Medical specialities and Department(s)

3. First results

The actual System Lists were configured especially for cardiology topics. The single diseases and procedures and the branches where filled with the common Pubmed Tags: Special Authornames Keywords, MeSH major topics, Title Words, Text Words and Journalnames.

For each user the system is able to submit a request computed from both the individual profile and the corresponding common profile. The whole profile can then be exported as XML-Meta-File. A user can see a weekly update of new abstracts form pubmed and internal ressources. A Grading on usefulness of the single papers by the user can then reconfigure the profiles depending on simple algorithms by adding single keywords or by adding negative lists of journals or keywords. At this point of time, the system gives just simple recommendations on that, the process itself is done by hand.

Figure 3 shows the architecture of the System.

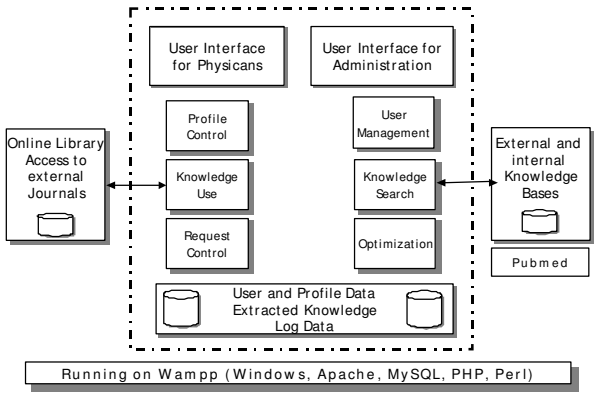

Figure 3. System architecture of the developed System

Results with the system with 230 doctors, 120 medical specialities and 800 Ilnesses/Procedures over a period of 4 weeks showed good results in the first runs, but only a little proportion of the users were “power users”. The semiautomatic update of profiles was done by exclusion of special keyword-patterns and association of keyword-patterns between similar users. As recommendations from the systems for profile-updates were not always optimal, it turned out, that more sophisticated algorithms are needed for better automatic optimisation of the complex user profiles and keywordpatterns.

4. Discussion

The developed system seems to be useful, as complex userprofiles can serve for knowledge update of new incoming files and publications. The approach to combine an automated notification-robot with an (semi)automated update and optimisation of user (search)profiles is new and can add economic benefit into the generation of searchpatterns and save time for the users on the mid-term run. The use of “Recommender Systems” algorithms could add a big portion of functionality to the system, as alterations of search profiles can be done automatic. These systems are under investigation and in practical use in diverse business scenarios for some years now [2]. As an example, Amazon.com uses the systems for book recommendations between online-customers [3].

The worth of the system may lay in the generation of profiles and the export and use over the described XML-Interface for other medical branches. This way benefit can be drawn form the generated search-patterns and they can be used more efficient. The Architecture is open for integration of the described “Recommender Systems”, which should be definitely topic of further investigation.

References

[1] Cartwright J, de Sylva S, Glasgow M, et al. Inaccessible information is useless information: addressing the knowledge gap. In: J Med Pract Manage (United States), 18(1), Jul-Aug 2002: p36-41.

[2] Landmann N. Einsatz und Potenziale von Recommender Systems im intranetbasierten Knowledge Management. Congress on Knowledge Management on 19.6.2002, Marburg University, Department for Mathematics and Informatics.

[3] See URL http://www.amazon.com

[4] See URL http://www.nlm.nih.gov/mesh/

Address for correspondence.

Christian Elsner
Thomasiusstr. 9
04109 Leipzig
Germany
Ch.Elsner@gmx.de