

Alternative interfaces for PubMed searches

Isabelle de Kaenel <isabelle.de-kaenel@chuv.ch>, Pablo Iriarte <pablo.iriarte@chuv.ch>

Medical Library ; Lausanne University Hospital - CHUV

Introduction

Over the last 5 years, encouraged by the NLM “[Licensee Research Use of NLM MEDLINE®/PubMed® Data](#)” and the public web services “[Entrez Programming Utilities](#)”, more than 30 new search interfaces for MEDLINE/PubMed have been made publicly accessible on the web. These developments, originated all over the world in academic medical centres, technical universities, bioinformatics departments, molecular biology laboratories, cancer research centres, government agencies such as the National Institutes of Health, and to a lesser extent in biotechnology and pharmaceutical companies.

Drawing on data mining, linguistics, statistical methods, knowledge management and new web technologies, specialists from different horizons teamed in order to extend and build upon many PubMed specificities and functions.

At the end of 2006, reference librarians at the Medical Library of the Lausanne University Hospital decided to monitor and analyse these new online tools, in order to decide whether they should be used as alternate search tools for specific information requests and even be introduced in training sessions for students and clinicians.

Method

To get an overview of these different retrieval systems, a literature search and a web search were conducted. The sites listed in the [NLM's registry of licensees who use MEDLINE/PubMed data for research purposes](#) and the MEDLINE interfaces section of the “[Metadatabase for the biological sciences](#)” were also systematically explored.

The background for the development of these new research tools became clearer. These applications are mainly designed to overcome some shortcomings and restrictions of the PubMed search interface : lack of proximity operators, some advanced functions that remain hidden - especially for many novice users - a rather dull display of long result lists, restricted personalisation, a lack of analytical tools to navigate through the results.

However, if the Entrez search engine, together with the size of the database, may produce challenges for the users, many researchers think that the enormous amount of PubMed data (millions of citations with free text abstracts) have a lot to offer if text mining methods were applied to parse the information contained in these references and abstracts (identification of emerging infectious diseases, discovery of gene-to-gene interactions and connection to diseases and/or existing drugs and support to drug or vaccine research).

To push the investigation further, more than a dozen of these MEDLINE interfaces were picked as the most mature and innovative applications. The search engines were classified according to specific features and function they incorporate.

Results

All these alternative search engines for MEDLINE/PubMed aim at enhancing interactivity and efficiency - either at the "input level" with user-friendly forms, assisted advanced queries, clear and effective search screens which help in building comprehensive and effective search strategies - or at

the output level with alternative sorting possibilities; original display and classification of the search results. These innovative functions help the users to evaluate and manipulate the references returned by a query; some of them try to work at both levels in order to help in building a precise search at one end, and, at the other end, to help in screening the set of references retrieved.

Below are some examples of functionalities that enhance user interaction with the database when searching :

- multilingual MeSH : [BabelMeSH](#) offers the possibilities to work with MeSH terms in 8 different languages
- guided search based on a PICO form : [PICO Linguist](#) displays a search form (available in 6 languages) that helps in building a structured clinical search based on the Patient / Intervention / Comparison / Outcome analysis
- free text, natural language query tool : [AskMEDLINE](#) is also aimed at answering clinical questions written in plain English
- similarity quest : with [Pubfinder](#) the starting point of a topic search is the input of a representative of at least 8 PMIDs. Based on this input a hit list of references is returned in order of relevance according to the algorithm developed by this search tool
- graphical search interfaces : the most striking example being [PubMed Interact](#) which presents slider bars to control search limits and parameters. This highly interactive interface pushes to the forefront the different traditional PubMed limits and filters

Examples of improvements in evaluating the relevance and sorting the set of references retrieved :

- relevance ranking : if relevance ranking exists in PubMed, it works on related articles only, not on the original query submitted. [ReleMed](#) is a new application which estimates a relevance score based on the query words
- graphical and interactive display of the extracted information : [AliBaba](#) provides a graphical view and offers advanced association mining
- generation of frequency tables : [Pubreminer](#) presents the results in the form of hyperlinked tables
- "on the fly" clusters of references grouped by topics derived from MeSH terminology: [ClusterMed](#) or [MESHPubMed](#) use the MeSH hierarchy to classify the results
- personal annotations and tagging of references : [Hubmed](#) offers a wealth of functions that allow the user to add personal information and links

Conclusion

The general conclusion is that these new search interfaces address many different needs and user profiles. As French speaking librarians, we were very attracted by the multilingual tools such as [BabelMeSH](#) and [PICO Linguist](#), knowing for example that clinicians working in private practice are still reluctant to plunge into large Anglo-Saxon databases without any language support.

Biologists and researchers that need to browse and scan vast amount of literature may be attracted by more sophisticated methods that mine information and bring out connections and associations hidden in long lists of abstracts.

As for the students they may be appealed by social tagging and web 2.0 technologies that they also use in recreational web sites.

In short, these new interfaces have to be introduced in training sessions, at least to appreciate the impact and interest they raise in the different groups of users.

These search tools certainly herald the transition from "resource centric" to "user centric" search systems.