PRODUCT REVIEW / ANALYSE DE PRODUITS

Product Review: PubMed PubReMiner

Product: PubMed PubReminer, developed by Jan Koster, Academisch Medisch Centrum, Universiteitvan Amsterdam

URL: http://bioinfo.amc.uva.nl/human-genetics/pubreminer/

Audience: PubMed searchers and those wishing to analyse PubMed search results

Purpose: A front-end that searches PubMed and produces frequency tables outlining the number of records published on a topic by year, journal, author, keyword, MeSH, substances and country

Compatibility: Web-based, open access. Plug-in available for Firefox and IE

Introduction

PubReMiner is one of a number of front-ends that analyse PubMed search results [1]. Each of the various interfaces has a specific purpose including simplifying PubMed searches, data-mining of PubMed search results, and facilitating access to full-text of articles retrieved in PubMed.

PubReMiner has been described previously in blog/wiki postings [2, 3] and has been compared to another leading front-end, Go PubMed [4]. It has also been used as a research tool in at least one study [5]. This review seeks to take a closer look at this resource, in particular commenting on its usefulness to both end-users and expert searchers.

How PubReMiner works

Initiating the search in PubReMiner is no different than it is in PubMed, i.e. the searcher enters keywords into a search box and clicks on a *Search* button. The search can be refined by limiting search terms to *Fieldtype* (*Title, Title+Abstract*, *Author*, *Journal*), *Publicationtype* (*All* or *Reviews*), and by *From/To Date*. The default limit for number of abstracts retrieved is 1000, but this can be changed to a maximum of 10,000.

Results from the initial search are reported in frequency tables for the following elements: *Year*, *Journal*, *Author*, *Word*, *MeSH*, *Substances,* and *Country*. The default for the *Author* display is to list all authors using their initials. The display can be adjusted to display only first authors, only last authors or to display the author's full first names. The default Word display lists all individual words included in the title, abstract, MeSH and registry name fields. This can be adjusted to display only title words or only title/abstract words. Country lists results by the country that appears in the PubMed Affiliation [AD] field.

Once the initial results are returned, terms can be selected from the various columns and added to the initial query as a means of refining the search and then, using the *GotoPubMed with query* button, run it directly in PubMed.

Ease of use

The initial search procedures are straight forward. Response time is acceptable. Larger retrievals will take longer to process; a pop-up provides an estimated processing time and a countdown to re-assure the searcher that the search engine hasn't stalled.

The initial search results can be 're-mined' by selecting terms from the frequency tables and adding them to the query box along with the terms originally entered. However, I found this to be a bit confusing at first. Based on the fact that AND is selected in the *Operator* pull-down menu, I assumed that terms selected from different columns would be ANDed together. However, this is not the case. Anything selected from the various columns will be ORed together and inserted into the query box and ANDed with what was entered originally.

For example, I entered *first nations diabetes* in the query box and browsed the resulting terms in each column. I selected *Indians, North American* and *Diabetes Mellitus, Type 2/Epidemiology* from the *MeSH* column and *Canada* from the *Country* column and clicked on *Search Again* hoping to retrieve results that discuss diabetes among First Nations individuals in Canada. The resulting search statement was:

> FIRST NATIONS DIABETES AND (%22CANADA %22 [AD] OR "DIABETES MELLITUS" [MH] OR "INDIANS, NORTH AMERICAN" [MH])

In order to retrieve the appropriate results, I had to manually adjust the search as follows in order to retrieve the intended results:

> (FIRST NATIONS OR "INDIANS, NORTH AMERICAN" [MH]) AND "DIABETES MELLITUS" [MH] AND CANADA.

Usefulness for Pubmed searching

One of the uses touted by the PubReMiner developer is to build efficient PubMed queries. While I believe the PubReMiner is helpful in assisting searchers in identifying

text words to incorporate into searches, PubReMiner is less useful in actually building the search without manual intervention from the searcher.

As the *First Nations diabetes* example above illustrates, using the re-mining feature requires a basic knowledge of Boolean logic to ensure that the results returned are appropriate. An understanding of PubMed field labels would also be helpful to ensure an effective search using the PubReMiner. For example, would the average end-user realize that the AD label indicates that the search engine is retrieving studies where the country appears in the address of the first author, rather than studies that discuss a particular country? If the latter is what is intended, the searcher must either delete the field label or select the country from the MeSH and/or Word tables.

The Word frequency table assists the searcher by listing the most frequently used text words contained in returned records. Rather than having to browse numerous abstracts one by one in PubMed, the searcher can quickly browse the PubReMiner Word table to identify relevant text words that can be incorporated into a search.

The MeSH table is useful for identifying the frequency of MeSH and MeSH/sub-heading combinations in a particular search. However, whether using the PubReMiner MeSH table is more useful than simply relying on PubMed's native automatic term mapping or using the PubMed MeSH database to select terms is debatable. For straightforward subject searches intended to identify the most relevant articles on topic, I see no benefit of using PubReMiner over the native PubMed interface.

Identifying authors and journals publishing in a research area

The availability in PubReMiner of Author frequency analysis of search results is immensely useful for those needing to identify the leading authors in a specific subject area for purposes such as identifying possible collaborators or subject experts to act as journal or grant peer reviewers. Likewise, those wishing to identify journals that frequently publish on a topic in order to select journals in which to publish will find PubReMiner to be a highly useful tool for this purpose. PubMed itself does not provide an option for listing results by author or journal frequency, nor do other major database vendors like Ovid or EBSCO.

Output

The frequency tables produced by PubReMiner (not the PubMed references themselves) can be saved as a text file. However, since it is strictly a text file, it has limited use. The ability to download PubReMiner results in an Excel-compatible format would be much more useful.

Available help and instructions

The on-screen help is minimal and trial and error will have to called upon to determine the best approach to searching and refining results and to determine what some of the search features actually do. For example, nowhere does it explain that the *Merge Similar Words* option is basically asking searchers whether or not they wish to truncate terms selected from the Word column. In addition, nowhere in the Help/FAQ does it mention that the same search conventions used in PubMed itself can be used in PubReMiner (e.g. field qualification can be used, double-quotes can be used to search exact phrases, PMIDs can be entered, etc.).

The PubReMiner developer appears to be highly responsive to user queries and suggestions. In addition to responding within an hour to a query about an error message, he also responded favourably to my comment that the instruction appearing above the initial search box, Enter Your PubMed Question, might lead searchers to believe they could enter their query in the form of a question. He agreed that this might be confusing and plans to change the wording upon the next interface update.

Conclusion

Expert searchers will find PubReMiner a valuable tool to assist in identifying relevant text words and MeSH terms, but end-users without a sound knowledge of search principles or PubMed search conventions will likely obtain better search results by searching PubMed directly. However, any individual – expert searcher or end-user – can easily make use of the *Author* and *Journal* frequency analysis tables whenever they need to identify experts in a subject area or journals that routinely publish studies on a particular topic.

References

1. PubMed Alternative Interfaces. 2012; Available at: http://hlwiki.slais.ubc.ca.login.ezproxy.library.ualberta.ca/index.php/PubMed_Alternative_Interfaces. [Accessed March 16, 2012].
2. Finding assigned MeSH terms and more: PubReMiner. Laika's MedLibLog. Retrieved from http://laikaspoetnik.wordpress.com/tag/pubreminer/
3. PubReMiner. HLWIKI Canada. Retrieved from http://hlwiki.slais.ubc.ca/index.php/PubReMiner.
4. Bradley S, Guistini D. Go PubMed vs PubReMiner for analyzing PubMed Results: A head-to-head comparison of two free web 'data-mining' tool. Presented at the CHLA/ABSC Conference, Calgary AB, May 26–30, 2011. Retrieved from http://chla-absc.ca/2011/sites/default/files/E_GoPubMed_vs_PubReMiner_Bradley_Giustini.pdf
5. Glynn R, Kerin M, Sweeney K. Authorship trends in the surgical literature. *Br J Surg*. 2010;97:1304–308.

Linda Slater
John W. Scott Health Sciences Library
University of Alberta
Edmonton, AB T6G 2R7
Email: linda.slater@ualberta.ca
Tel: 780-492-7948
Fax: 780-492-6960