An Open-Source, Mobile-Friendly Search Engine for Public Medical Knowledge

Matthias SAMWALD [a,1] and Allan HANBURY [b]

[a] *Section for Medical Expert and Knowledge-Based Systems; Center for Medical Statistics, Informatics, and Intelligent Systems; Medical University of Vienna, Austria*
[b] *Information & Software Engineering Group; Institute of Software Technology and Interactive Systems; Vienna University of Technology, Austria*

Abstract. The World Wide Web has become an important source of information for medical practitioners. To complement the capabilities of currently available web search engines we developed FindMeEvidence, an open-source, mobile-friendly medical search engine. In a preliminary evaluation, the quality of results from FindMeEvidence proved to be competitive with those from TRIP Database, an established, closed-source search engine for evidence-based medicine.

Keywords. Information retrieval; medical care; World Wide Web; clinical decision support

Introduction

It has been shown that web-based search engines such as Google can aid the diagnostic process (1) and that the availability of an online information retrieval system increased the percentage of correctly answered medical questions from 21% to 50% in a group of 75 clinicians (2). In a recent study we conducted, medical professionals were shown to frequently use general-purpose search engines such as Google, medical research databases and – perhaps surprisingly – Wikipedia to answer medical questions online (3). A potential problem with these resources is that most of them either return large amounts of clinically irrelevant or untrustworthy content (e.g., Google), or that they are mainly focused on primary scientific literature that makes selection of clinically relevant publications very time-consuming (e.g., PubMed).

Here we report on our development of FindMeEvidence, an open-source, mobile-friendly web search engine optimized for medical information needs. The goal of the FindMeEvidence project is to improve efficient access to medical evidence on the web by providing a free, easily customizable, light-weight solution for medical information retrieval.

---

[1] Corresponding Author.

1. Methods

FindMeEvidence was built around the open-source Apache Solr 4.4. information retrieval system. Server-side scripts were written in PHP 5. An overview of the system architecture is shown in Figure 1.
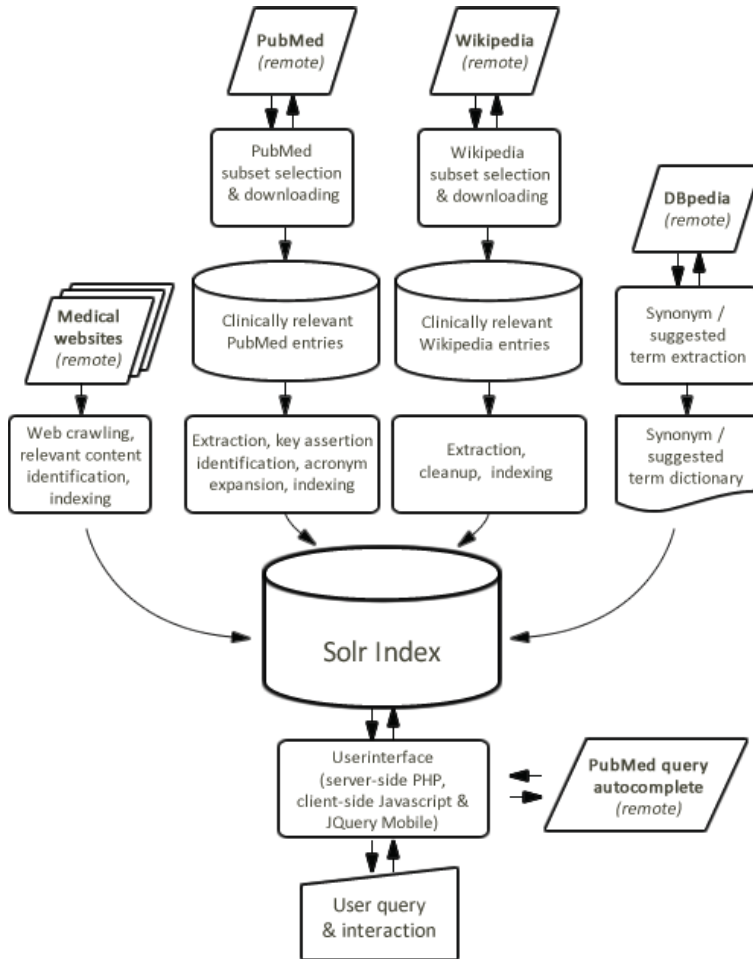

Figure 1: FindMeEvidence system components

The selection of medical data sources was based on scientific studies of digital information needs of medical professionals (3), as well as a review of popular and trustworthy web resources providing medical evidence. Only datasets in English were included. Based on these criteria, the following datasets were integrated into FindMeEvidence: A clinically relevant subset of PubMed and Wikipedia, Merck Manual (4), Medscape (5), NICE Clinical Knowledge Summaries (6), Guideline.gov (7), BestBETs (8) and ATTRACT (9). Wikipedia was among the included datasets because it was reported as an important tool for medical professionals (3). Wikipedia can be used to get a quick overview of a novel subject matter as well as to retrieve term

definitions (10), even though the quality of detailed medical and pharmacological information in Wikipedia was found to be modest (11).

One author (MS) conducted a simple, preliminary evaluation of the quality of search results returned by the system based on a list of 36 medical queries that were derived from a 24-hour sample of the query logs of the PubMed search engine (12) and a sample query log of the TRIP database. The queries consisted of simple Boolean queries with a length of one to five words. For each query, the likely information need behind the query was written down. Then, it was checked whether FindMeEvidence and TRIP database (as a control) returned relevant information within the first 5 query results each, without using any result filters available in the search engines.

2. Results

The search engine is accessible at http://FindMeEvidence.org/, the source code repository is accessible at (13). Figure 2 shows a screenshot of the user interface of the software.
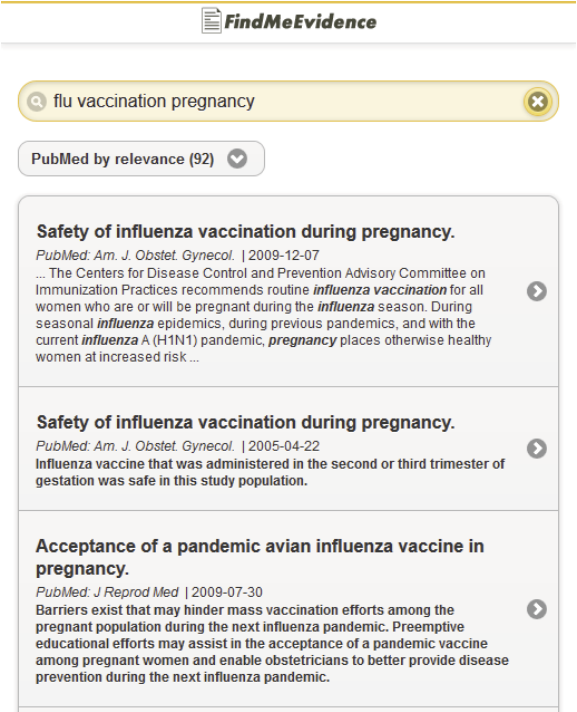

Figure 2: Screenshot of a FindMeEvidence result list containing clinically relevant research results. In the second and third result, key findings were identified in the abstracts and are shown in the result previews.

Out of all test queries (N = 36), FindMeEvidence results met success criteria for 25 (69,4%) of the queries, while TRIP Database results met criteria for 17 (47,2%) of the queries. It should be noted that both the results of FindMeEvidence and TRIP Database could have been further improved by making use of their result filter functionalities –

this is especially relevant for TRIP Database, which offers sophisticaed functionalities for result filtering based on different types of evidence.

3. Discussion

The first public release of FindMeEvidence we present here was able to provide useful answers to a variety of realistic medical queries and could compete with systems that are the state-of-the-art in medical information search. The evaluation results allow us to further optimize data source selection and ranking. A common reason for failure of both FindMeEvidence and TRIP database was that for some general queries, very specific results – such as results from very specific clinical studies reported in PubMed abstracts – were returned first. Although users can potentially exclude these specific results by selecting an appropriate search facet, this problem should be better addressed in future versions of the software.

An objective evaluation of modern search engines is very difficult, since many relevant aspects of the overall user experience cannot be captured by the simple assessment of precision and recall of document retrieval (14). Furthermore, the needs of different medical user groups vary significantly, and different search engines might fit the preferences of different groups. An important goal of FindMeEvidence is to complement the qualities of existing solutions by providing a light-weight, focused entry point to the most relevant openly available medical content.

Since FindMeEvidence is open-source software, local institutions can easily create their own versions of the software that are tailored to their specific needs, such as including descriptions of institutional standard operating procedures.

A major next step in the evaluation and further improvement of FindMeEvidence is to test it with medical professionals in their daily routine. The preliminary evaluation results presented in this paper suggest that FindMeEvidence has already reached a level of maturity that makes such an evaluation feasible.

3.1. Related Work

Several systems geared towards providing medical information retrieval capabilities in clinical settings have been described in the literature. To our knowledge, FindMeEvidence is the only currently available system that is open-source, built on an industry-strength information retrieval engine, and optimized for cross-platform and mobile web access. Some examples of related systems are listed below.

The TRIP Database (15) is a publicly available, closed-source search engine for evidence-based medicine and was used as a comparison for our evaluation. TRIP Database has a large user base and has processed over 100 million user queries.

One of the earliest systems with goals comparable to FindMeEvidence was Quick Clinical, a federated medical search system developed by Coiera *et al.* (16). Quick Clinical introduced sophisticated meta-search filters that allow users to optimize query results by explicitly stating the types of their information needs (e.g., diagnosis, etiology) in addition to a free text query. Quick Clinical was extensively evaluated and was shown to improve clinical question answering (2). To our best knowledge, the Quick Clinical system is currently not publicly available.

Khresmoi (17) is a large European research project developing a medical search engine platform around the GATE Framework and semantic technologies. Compared to

FindMeEvidence, the Khresmoi system is more geared towards large-scale data processing and the support of machine translation, and its system requirements are significantly higher.

3.2. Conclusion

FindMeEvidence demonstrated its potential to become a useful addition to the digital toolset available to medical professionals. Still, there is a lot of room for potential improvements and we invite interested parties to contact us and join the future development and dissemination of the system.

Acknowledgements

References

[1]   Tang H, Ng JHK. Googling for a diagnosis—use of Google as a diagnostic aid: internet based study. BMJ. 2006 Dec 2;333(7579):1143 –1145.
[2]   Westbrook JI, Coiera EW, Gosling AS. Do Online Information Retrieval Systems Help Experienced Clinicians Answer Clinical Questions? J Am Med Inform Assoc JAMIA. 2005;12(3):315–21.
[3]   Kritz M, Gschwandtner M, Stefanov V, Hanbury A, Samwald M. Utilization and Perceived Problems of Online Medical Resources and Search Tools Among Different Groups of European Physicians. J Med Internet Res. 2013;15(6):e122.
[4]   THE MERCK MANUALS - Trusted Medical and Scientific Information [Internet]. [cited 2013 Oct 15]. Available from: http://www.merckmanuals.com/
[5]   Diseases & Conditions - Medscape Reference [Internet]. [cited 2013 Oct 15]. Available from: http://emedicine.medscape.com/
[6]   NICE Clinical Knowledge Summaries (CKS) [Internet]. [cited 2013 Oct 15]. Available from: http://cks.nice.org.uk/
[7]   National Guideline Clearinghouse | Home [Internet]. [cited 2013 Oct 15]. Available from: http://www.guideline.gov/
[8]   Welcome to BestBETs [Internet]. [cited 2013 Oct 15]. Available from: http://bestbets.org/
[9]   ATTRACT | Home [Internet]. [cited 2013 Oct 15]. Available from: http://www.attract.wales.nhs.uk/
[10] Heilman JM, Kemmann E, Bonert M, Chatterjee A, Ragar B, Beards GM, et al. Wikipedia: a key tool for global public health promotion. J Med Internet Res. 2011;13(1):e14.
[11] Clauson KA, Polen HH, Boulos MNK, Dzenowagis JH. Scope, Completeness, and Accuracy of Drug Information in Wikipedia. Ann Pharmacother. 2008 Jan 12;42(12):1814–21.
[12] Herskovic JR, Tanaka LY, Hersh W, Bernstam EV. A Day in the Life of PubMed: Analysis of a Typical Day's Query Log. J Am Med Inform Assoc JAMIA. 2007;14(2):212–20.
[13] bricoleur-fast-medical-search - Building a highly efficient search engine for medical professionals - Google Project Hosting [Internet]. [cited 2013 Nov 26]. Available from: https://code.google.com/p/bricoleur-fast-medical-search/
[14] Jouis C, Biskri I, Ganascia J-G, Roux M, editors. Next Generation Search Engines: Advanced Models for Information Retrieval [Internet]. IGI Global; 2012 [cited 2013 Oct 28]. Available from: http://www.igi-global.com/chapter/framework-evaluating-retrieval-effectiveness-search/64437
[15] Trip Database [Internet]. [cited 2013 Oct 15]. Available from: http://www.tripdatabase.com/
[16] Coiera E, Walther M, Nguyen K, Lovell NH. Architecture for Knowledge-Based and Federated Search of Online Clinical Evidence. J Med Internet Res. 2005 Oct 24;7(5):e52.
[17] Khresmoi - Medical Information Analysis and Retrieval [Internet]. [cited 2010 Dec 6]. Available from: http://khresmoi.eu/