Reviewer's report

Title: CDAPubmed: a browser extension to retrieve EHR-based biomedical literature

Version: 1 Date: 28 March 2011

Reviewer: Zhiyong Lu

Reviewer's report:

CDAPubMed would be a nice addition to the PubMed family since it offers an extension to integrate with EHR data to facilitate more focused PubMed searches. As a result it has the potential to benefit those who work in the related field. The writing is clear and mostly easy to follow.

Major:

My main concern is the use of the word “relevance” in the sections of Query Generation and Results. In my view, adding MeSH terms in each iteration makes the query more specific / focused, but this does not automatically translate to more relevant results. I agree this strategy should help retrieval in general, but to warrant the use of different levels of relevance (as discussed in Results), more experiments with human relevance judgments on a large set of retrieval data would be necessary. The authors should also discuss (a possible weakness) that adding MeSH terms to the query in each iteration favors precision, but at the same time it may also result in a loss of some relevant publications.

Another important discussion to be included (maybe for future work) is how the sequence of adding MeSH terms to a query may affect final results given there may be a list of MeSH terms available in an EHR. For a physician with limited time and perhaps modest knowledge about Boolean queries, it would be nice to prioritize MeSH terms to be added in a meaningful order. This could also avoid getting zero results when a user randomly combines two disjoint MeSH terms from such a list.

Minor:

1. The authors could take advantage of MeSH major topics (http://www.nlm.nih.gov/bsd/disted/mesh/major.html) to further differentiate articles with the same “relevance” score.
2. Please be consistent about PubMed in itself in and in CDAPubMed. Both PubMed and Pubmed are seen in the paper.
3. Avoid use of dash when it is not necessary. E.g. page 2. Last sentence in second paragraph of Background. “—e.g. Google, Yahoo, or Bing.”
4. PubMed-EX (Tsai et al., Bioinformatics, 2009) is a web browser extension. So it should be relevant to the discussion in page 6.
5. Page 2. 'such HubMed' should be “such as HubMed”
6. Page 4. “a NLP” should be “an NLP”
7. Page 6. “a tool that improved” better be “a tool that can improve”
8. Figure 3 is not very legible. Please improve its resolution.

Finally, there was no link to the software in the submission so I was not able to do an assessment on the software.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests