

## Reviewer's report

**Title:** AnonMed: Sentence-level search engine for the MEDLINE database of biomedical articles

**Version:** 2 **Date:** 10 May 2006

**Reviewer:** Stefan J Darmoni

### Reviewer's report:

#### General

Very interesting paper with an original idea to improve information retrieval in one major biomedical database (MEDLINE). Any MEDLINE end-user has the experience to read a lot of false positive records for a query. Therefore, the main idea of this paper is quite exciting... at least for most of the users. It is surely not the case for exhaustive searches...

The Results section and the Table are clear.

Should be accepted after the following major compulsory revisions

-----  
Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

Most of the readers (even in medical informatics) know quite well the definition of sensitivity and sensibility. Reduce this part in the Background Section.

But in information science, other metrics are used: mostly precision and recall.

To have more details on information science metrics, look at Meystre S, Haug PJ.

Natural language processing to extract medical problems from electronic clinical documents: Performance evaluation.

J Biomed Inform. 2005 Dec 5; [Epub ahead of print]

Do not confuse in page 1 PubMed which is mainly an interface and MEDLINE which is a bibliographic database. Please define both more properly.

The evaluation method need to me more precisely described. Are the experts really blind in this study? it is not really clear.

Furthermore, in the Evaluation method section, I will expect to read more clearly how the authors have choosed the two examples. Why they have choosed only two?

In the Discussion section, I will expect to read about two major trends of information retrieval in medicine.

1/ using matrix of UMLS co-occurences

2/ using UMLS semantic network.

I really think that the authors need to compare their scientific hypothesis and results with these two major tracks of this discipline.

-----  
Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

I will expect at least a reference for the first sentence of the Discussion.

I am quite sure that the NLM are able to provide quite precise figures about it, to avoid: 'it seems...'

-----  
Discretionary Revisions (which the author can choose to ignore)

It is a more general comment. The idea mentioned in the Discussion section about searching only in the

Title metadata is quite interesting. We have implemented it in our Web site [www.cismef.org](http://www.cismef.org) which is a major difference with the PubMed Web site.

**What next?:** Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

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

**Quality of written English:** Acceptable

**Statistical review:** Yes

**Declaration of competing interests:**

I declare that I have no competing interests' below