Reviewer's report

Title: Automated Systems to Identify Relevant Documents in Product Risk Management

Version: 1 Date: 16 August 2011

Reviewer: Martijn Schuemie

Reviewer's report:

In this manuscript, the authors employ text-mining to automatically distinguish between useful and non-useful articles for drug safety screening. Off-the-shelf text-mining tools are used, but this is fine. In general, the article is well-written, and methods used appear sound. However, several changes must be made before the manuscript is suitable for publication:

Major Compulsory Revisions

1. First and foremost, the description of the field of text-mining as included in the introduction is not at all representative. This is a field with many publications over the years, and the authors seem to have picked a few at random. I suggest referring to some review articles instead (a simple search in Pubmed already found http://www.ncbi.nlm.nih.gov/pubmed/21245076, but many others do exist), and there are several articles that are especially relevant for the work described here, such as http://www.ncbi.nlm.nih.gov/pubmed/19376821, where also PubMed records are classified. Also, the authors mix up the broad field of text-mining with the more narrow sub-field of literature based knowledge discovery (starting with the Swanson paper cited by the authors). The work described in this manuscript falls in the category of text-mining, but has nothing to do with literature based knowledge discovery.

2. The authors compare their method with a random ranking of the papers retrieved by a query, but this is like comparing a new drug to placebo when alternative therapies are available. The baseline should be a standard information retrieval system, for instance one that ranks documents based on TFIDF in relation to the search query. This could be a standard instance of Lucene, for example.

Minor Essential Revisions

p4. "such as Therapeutic Goods Administration in Australia, Food and Drug Administration in the United States of America and European Medicines Agency in Europe." should be replaced with "such as the Therapeutic Goods Administration in Australia, the Food and Drug Administration in the United States of America and the European Medicines Agency in Europe."

p8, figure 1: please include the 'generalizability set' in figure 1, so figure 1 depicts
all data sets used

p9. Please explain why synonyms from Omniviz were used instead of more obvious sources such as UMLS

p19. I don’t understand why spelling errors would be a bigger problem in translated abstracts (translators are often very good in English). Have the authors observed these problems, or is it just speculation?

Tables 1 and 3: please specify in the figure caption whether the generic or the specific features were used.

Discretionary Revisions

To be honest, I'm surprised that including IDF improves performance. The IDF is a constant value applied to a feature over all documents. Could the authors give an explanation why this increases the ability to distinguish between documents?

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