Reviewer's report

Title: Mining FDA Drug Labels for Medical Conditions

Version: 2 Date: 9 January 2013

Reviewer: Mathaeus Dejori

Reviewer's report:

Major Compulsory Revisions

1. The paper describes a system to extract drug-medical condition from FDA drug labels with the use cases of populating ADRs (adverse drug reactions). The use case is highly compelling as drug labels seem to be an ubiquitous source of information.

2. The key contribution of this paper is to apply CRF (Mallet) as an NER tool and to feed it with features produced by cTAKES to detect medical conditions. Extracting medical conditions from free text has been studied extensively and is part of many off-the-shelf packages. Extracting medical conditions in the context of ADRs using supervised methods has been proposed as well by Aaramik et al (cited by the authors) and therefore lacks novelty.

3. While the methods used are off-the-shelf (cTAKES, Mallet) the experimental design is thorough, comparing the proposed approach with two rule-based systems. However, given the reported Precision/Recall measures on partial/exact matches such a system won't be ready to be used. For instance a Recall of .88 on partial matches means somebody would still have to manually go through all of the drug labels to capture the remaining 12%.

4. I would like to see an ROC analysis that investigates how much Precision will decrease with increasing Recall.

5. What was the inter annotator agreement?

6. (When) will the data made public?

Minor Essential Revisions

some abbreviations are not spelled out initially nor listed in the appendix. For example, cTakes on page 6.

Some sentences are difficult to understand, e.g. page 11 "The findings on the token level are the tokens of the documents."

Discretionary Revisions
- none

Level of interest: An article whose findings are important to those with closely
related research interests

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

I declare that I have no competing interests