Reviewer's report

Title: A Controlled Trial of Automated Classification of Negation from Clinical Notes

Version: 2 Date: 18 January 2005

Reviewer: Jeremy E Rogers

Reviewer's report:

General
The central section of this paper, reporting the original work, is much improved but the introduction, although now more relevant to what follows in so far as it reviews some prior NLP work, needs alteration.

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

The introduction needs greater focuss. The first paragraph of page 2 seems a complete non-sequitur from the preceding paragraph, and the highly abbreviated history of computational linguistics that follows is unnecessary. In my opinion the first four paras of page two should be condensed into a single paragraph, or even deleted altogether. It is sufficient that the references 1 thru 10 of prior NLP work are cited in the first paragraph.

The second sentence in para 6 on page two is a duplicate of the final paragraph on the same page.

The results section still doesn't make it clear whether the 14,792 healthcare concepts identified was the count of the total number of different clinical concept classes encountered (but where more than one instance of a given concept might appear across the 41 texts) or the count of the number of instances of a healthcare concept (but where the number of different classes of concept encountered was somewhat less than 14,792). Out of all the 300k or more SNOMED codes available, how many were actually used at least once in order to express these 14,792 concepts?

My previous review asked that the headline figures for precision and recall given in the results section of the manuscript abstract should be qualified. For example, average sensitivity 97.2% (range 50% - 100% across identified document subsections) average specificity 98.8% (range 33% - 100%). The statement that clinicians are happy with 95% or greater accuracy must be reconciled with the much lower accuracies reported for some sections of the documents analysed.

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

The use of the terms 'clinical record', 'medical record' and 'records' in the introductory paragraphs of the study design and results sections remains ambiguous. I would normally understand these terms to mean the entire corpus of clinical texts relating to one patient - including every clinic letter ever written about that patient - and comprising 100 or more discrete documents per patient, whereas within this manuscript these terms instead usually mean one instance of one particular type of document (unique to the Mayo) within such a corpus, for one patient.

Perhaps the term 'clinical evaluation report document' would be less potentially confusing than
'clinical record'. Similarly, where the term 'record' is used alone, I suggest substituting either 'report' or 'document' when this refers to one document and not all documents for one patient.

Discretionary Revisions (which the author can choose to ignore)

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

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

**Quality of written English:** Acceptable

**Statistical review:** No