Summary

This paper is about using a statistical machine translation tool (IBM model 1) to predict keywords for documents given the title of the document as a source.

Comments and Criticisms

First off, the paper is reasonably well-written; however, it would require the kind editing hand of an individual more familiar with the grammatical rules of English article use before publication. Examples of missing articles abound in the text, and, though distracting, didn’t substantially impact the import of this review.

The introduction to this paper is relatively long-winded and unnecessarily vague. It could be shortened by half and the paper would not be substantially impacted.

My first serious criticism with the work is a claim made in the last paragraph of the first page: namely, that the method in the paper is unsupervised. The word unsupervised applies to two things simultaneously here and both of them are wrong. The proposed method is supervised in the sense that human editors must have generated titles for the training corpus \_and\_ the statistical technique is supervised in the sense that the conditional distribution of keywords given source is learned rather than the joint distribution. So, the claim of this being an unsupervised approach seems rather too strong.

The notation used in the related work section, particularly section 3, is confusing and imprecise. In particular <D,T> in the algorithm box is extremely unclear. Is this an expectation or a set? Is D still a large collection of documents? Of so, what is T? A large collection of titles? It would help readers tremendously to very carefully lay out notation once: what is a keyword? Is it a collection of words? Is it ordered? What is a word? Is it a type or a token? Sometimes capitals are sets. Sometimes small type is a set, a sequence, a value, a token… notation is too imprecise.

Although TFIDF is familiar to a large number of people, it is still an acronym and still needs to be calculated. It should be explained, particularly since in many places in the paper the set of TFIDF weights are used as a probability distribution. Do you normalize these weights in some way so that this use is valid?

The choice of \delta seems arbitrary. Given this method of choosing translation pairs, the model and method can’t be used for prediction in a test document because there would be no way of selecting the most “semantically similar” sentence to the keywords, because, of course, the keywords would be unknown.

Footnote 1 should have a reference for the claim.

Why do you have to translate in both directions, as in Eqn. 3? This isn’t explained clearly enough.

The notation in section 3.3 is, again, confusing. What is p? What is d?

In section 4.1 you take away the first part of the title of the paper “keyword generation” and replace it with “keyword extraction.” Shouldn’t the title be a little less general too then?

For SMT, doesn’t setting the harmonic factor \lambda = 1 mean that you aren’t actually taking a harmonic mean? If so, then why lay it out that way in the first place?

Then, in the experiments, it appears that you are doing keyword generation/extraction from the titles of documents, not the documents themselves as claimed.

Relevance 3

Significance 2

Soundness 6

Originiality 3

Evaluation and comparison 5

Clarity 4

overall 3