Text Mining Exercise week 5 report

Guido Zuidhof, s4160703

1.

The selected "best answer" is the correct one. Source:

> The candidate answer set for a given question is composed by one positive example, i.e., its corresponding best answer, and as negative examples all the other answers retrieved in the top N by the retrieval component.

Although I must say it's not entirely clear. Elsewhere they note that they use no metadata and only the textual content:

> Again, since our interest is in investigating the utility of the answer textual content, we use only information extracted from the answer text when learning the scoring function.

2.

The learning-to-rank approach used is pairwise (and it's called - or based on - Perceptron).

3.

The question reads how they bridge the gap *without* using external knowledge bases such as WordNet. From the way I read it in the paper, they do in fact use external knowledge bases (they normalize all words to their WordNet lemmas).

However, they do use additional features which does not depend on these external knowledge bases. Namely, the **Translation Features**. This uses maximum likelihood estimation in some Bayesian framework to determine the probability that some question is a translation of an answer. Also, they use the **Web Correlation** features, which likely contributes to briding this gap as well.

4.

The share of result sets that have the (single) correct answer in the top N results. I suppose the term is a reasonable descriptor for this definition, although the ordinary definition of recall has something to do with the share of useful answers and as N increases this would decrease. But this Recall@N does not have this characteristic. The slides call this **Success@N**, which I think is more correct.

5.

I think the feature set used in the paper would be applicable to a reasonable extent. I think the **Density and Frequency Features** would not be useful, as these measures would really only work for longer answers, whereas the answer of factoid questions are generally quite short. I am not an expert, but as for the other features I can not see why they would not be useful to have.