

# Assignment D – On Paper

Eirik Isene

April 4, 2016

## 1 Query term proximity

A scoring method that considers the relative closeness of the query terms is generally desirable. How can it be integrated in a search engine? Provide a few alternative options for such implementation, and describe the advantages and disadvantages of each.

### 1.1 Suggestion A

We can implement a scoring that takes into account the closeness by dividing the document length with the distance between the query term in the document. This value can then be multiplied by the score of the document, this will in turn give high scores to the documents in which the terms are close, and lower scores to the documents with a long distance between. One of the biggest disadvantages of this is that documents that differ greatly in size will get uneven results, so longer documents will get higher values etc. This can be taken into consideration by scaling this score logarithmically, then great values won't have such a big impact.

### 1.2 Suggestion B

Another way to do this would be to add a maximum distance, and only consider results where the query terms are within this range of each other in the document. A great disadvantage to this method is that we might be trimming away a lot of useful results because one or more of the words are outside of this range, but don't really matter much to the query.