

AALTO UNIVERSITY

T-75.4400 INFORMATION RETRIEVAL

ASSIGNMENT 2

Comparing ranking methods for information retrieval

Antti PARTANEN
antti.partanen@aalto.fi
295967

Vikram KAMATH
vikram.kamath@aalto.fi
XXXXXX

April 21, 2015



Aalto University
School of Science

Too messy titlepage?

Abstract

In this paper, we compare compare two ranking methods: VSM (Vector Space Model) and BM25 for document collection from ACM digital library database. Furthermore, we study the effects of morphological techniques, including stemmers and stop words.

Abstract maybe not necessary? Overlap with introduction?

1 Introduction

The task of Information retrieval is to obtain information resources relevant to the imminent need from information resources. Resources include books, journals and other documents. Searches can be based on meta-data or on full-text indexing.

Automated information retrieval systems are used to reduce the so called *information overload*, i.e. the difficulty of a person to understand an issue due to overwhelming amounts of information. Most existing text retrieval techniques rely on indexing key-words. Unfortunately keywords or index terms alone cannot adequately capture the document contents, resulting in poor retrieval performance. Yet, keyword indexing is widely used in commercial systems because it is still the most viable way by far to process large amount of text.

Efficient and effective information retrieval techniques are critical in managing the increased amount of textual information available in electronic form. Therefore, it is no wonder that currently the most visible IR systems are web search engines.

In this paper, we compare the two information retrieval techniques: Vector Space Model (VSM) and BM25. Furthermore, we analyze what effects does the usage of stemmers (porter stemmer) and stopwords have with the retrieval result. The evaluation of our experiments is visualized with eleven point precision recall curves.

Draft version: aka needs modifying + references

2 Techniques

Techniques that we used were VSM and BM25.

Add text and references

2.1 VSM

2.2 BM25

2.3 Stemmers

2.4 Stopwords

3 Evaluation

Add text and charts

Precision

Recall

Eleven point precision recall curve

4 Conclusions

Add conclusions

5 Contributions

Mind map style of work division might look neat?

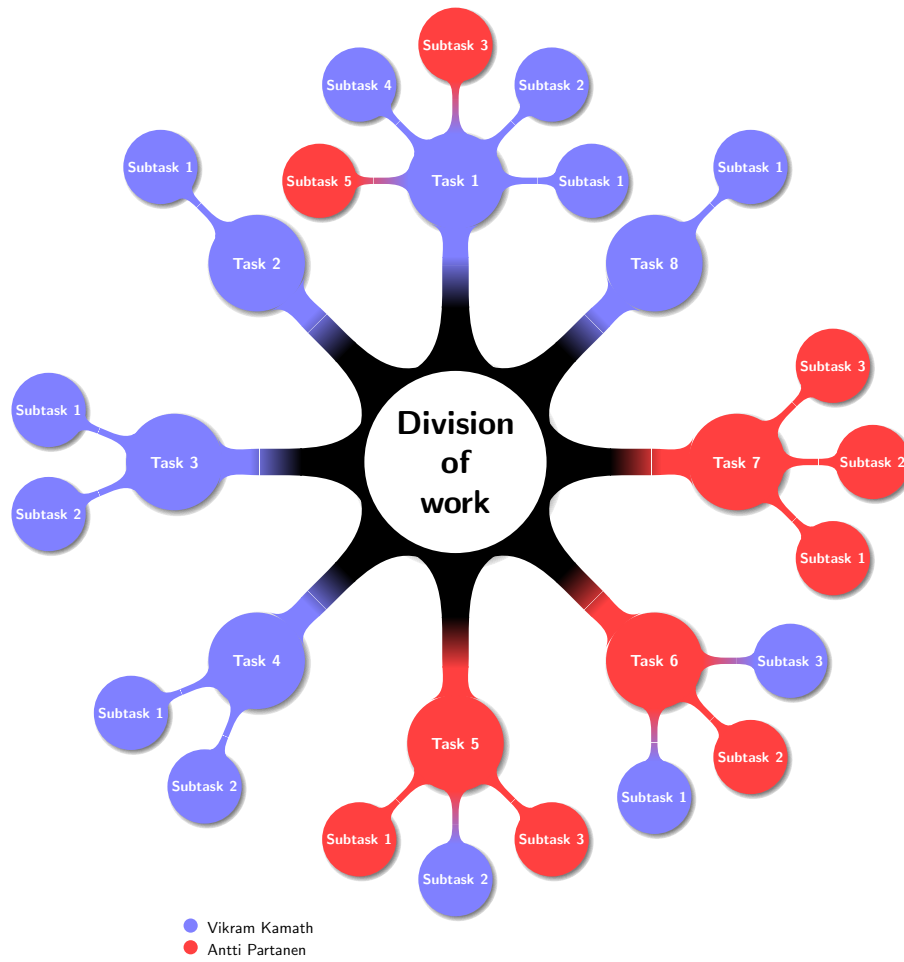


Figure 1: Contributions represented in a mindmap

References

- [1] WikiBooks. \LaTeX . <http://en.wikibooks.org/wiki/LaTeX/>, 2008.

Add more references

A Installation instructions

Add instructions

Here is a brief systems installation and system configuration instructions.

1. ...
2. ...