

# Foundations of Information Retrieval

## Exercise 5

Exercise session: 27.11.2014

Questions to: Helge Holzmann, holzmann@L3S.de

### 1. TF/IDF weighting and the Vector Space Model

Given is a document collection with two documents:

**Document 1:**

*A good cook could cook as much cookies as a good cook who could cook cookies.*

**Document 2:**

*Best chocolate chip cookies recipe.*

- a. Index the both documents using an inverted index. Apply case folding, use stop words list: {a, as}. Include TF and DF values at suitable positions in the index.
- b. Compute the cosine similarity between these documents and the following queries. Apply the **l<sub>tc</sub>** weighting from the lecture.

$Q_1 = \text{cookies}$

$Q_2 = \text{how to cook chocolate cake}$

- c. Explain the similarity scores! How would the result change with the **l<sub>nc</sub>** weighting?