ifis
Institut für Informationssysteme
Technische Universität Braunschweig

Information Retrieval and Web Search Engines
Summer Semester 2011

Prof. Dr. Wolf-Tilo Balke and Joachim Selke

Homework Assignment 2

Exercise 2.1

What is the relation between the Boolean retrieval model and the fuzzy retrieval model? Discuss the major similarities and differences.

Exercise 2.2

What are possible problems of using the Jaccard index for measuring term similarity? Give an example of two rather dissimilar terms that would typically yield a high Jaccard index.

Exercise 2.3

What is the basic idea underlying Ogawa's approach to deriving fuzzy term weights? (Do not use any formulas in your answer!)

Exercise 2.4

Given a document collection that is stored on disk using an inverted index (with a term weight assigned to each term–document pair). What is the computational complexity of calculating the cosine similarity between two documents?

Exercise 2.5

What is the purpose of normalizing a document's vector representation for document length?

Exercise 2.6

What is the basic idea underlying the TF–IDF weighting scheme? Why should we care about how often a term occurs in the collection? Give an illustrating example of your own.

Exercise 2.7

What is the difference between prior and posterior probability? How are both related to Bayes' Theorem?