

Homework Assignment 2

Homework Assignment 2 Please note that even though the homework assignments are optional, you're still highly encouraged to answer them, as they will help you prepare for your final exam. Solutions can be sent by email at ghammad@ifis.cs.tu-bs.de or can be given back to us at the lecture.

Exercise 1: Boolean Model

Given the following document collection:

- D1 = {t1, t3, t4, t5}
- D2 = {t1, t5, t3}
- D3 = {t1, t2}
- D4 = {t3, t5, t4}

1.1) create an invert index for all the terms of the vocabulary

1.2) perform the query Q1: *(t1 and t3) but not (t4 or t2)* over the inverted index (**hint use De Morgan's Law: "not (A and B)" is the same as "(not A) or (not B)"**)

1.3) convert the query Q2: *t1 and ((t3 and not t2) or t4)* into a conjunctive normal form and perform it over the inverted index

Exercise 2: Fuzzy Retrieval Model

Given the following fuzzy weights for D1 = {t1/0.4, t3/0.8, t4/0.8, t5/0.2, t2/0.1}

2.2) compute the relevance of the D1 to query Q2

2.3) what is the motivation behind using fuzzy weights?

2.4) what are the main differences between the fuzzy retrieval model and the Boolean retrieval model?

Exercise 3: Vector Space Model

3.1) why doing normalization in the vector space model makes sense? Is it always a good idea?

3.2) what is the idea of TF-IDF? What motivates its use?

3.3) explain the assumptions of the vector space model that do not hold in general.