



HW2: More Information Retrieval Model

Please note: Even though the homework assignments are optional **and will be neither corrected nor graded**, you are encouraged to answer them, as they will help you prepare for your final exam. Send your homework by email to: usman@ifis.cs.tu-bs.de. Please write your solutions in English. **Due date:** before the next lecture. Please do not forget your **Matrikelnummer** and your **full name** on your solutions.

1. Given the following subset of documents from a document collection:

$$\begin{array}{ll} D_1 = \{t_1, t_5, t_9\} & D_4 = \{t_4, t_5, t_{10}\} \\ D_2 = \{t_1, t_2, t_4, t_5, t_9\} & D_5 = \{t_3, t_5, t_6, t_7\} \\ D_3 = \{t_3, t_6, t_7, t_8\} & D_6 = \{t_1, t_2, t_{10}\} \end{array}$$

- Create an inverted index for the following terms: t_1, t_2, t_3, t_5, t_8
- According to the Boolean model evaluate the following queries:

- $q_1 = (t_1 \text{ or } t_5) \text{ but not } (t_3 \text{ or } t_2)$
- $q_2 = (t_1 \text{ and } t_5) \text{ or } (t_3 \text{ and } t_2)$

2. Given is the following subset of documents from a document collection and following queries:

$$\begin{array}{ll} D_1 = \{t_1, t_5, t_9\} & D_4 = \{t_4, t_5, t_{10}\} \\ D_2 = \{t_1, t_2, t_4, t_5, t_9\} & D_5 = \{t_3, t_5, t_6, t_7\} \\ D_3 = \{t_3, t_6, t_7, t_8\} & D_6 = \{t_1, t_2, t_{10}\} \end{array}$$

- $q_1 = (t_1 \text{ or } t_5) \text{ but not } (t_3 \text{ or } t_2)$
- $q_2 = (t_1 \text{ and } t_5) \text{ or } (t_3 \text{ and } t_2)$

a. (Fuzzy Logic Model)

Given the following fuzzy weights for $D_1 = \{t_1/0.4, t_2/0.8, t_3/0.8, t_5/0.2, t_8/0.9, t_9/0.6\}$

- Compute the relevance of document D_1 to both queries q_1 and q_2
- What is the motivation behind using fuzzy weights?
- Assuming the following journal collection: D_1, D_4 and $D_6 = \{t_5, t_9\}$ constitute an entire document collection, compute the fuzzy membership degree for the collection's terms with respect to D_1 .

b. (Coordination Level Matching)

- Given the following query: $q = \{t_1, t_2, t_4\}$ on the document collection defined in 1.c return level-based matching of these documents for q .