

Information Retrieval

Teilleistung 1

Dmitri Bajew, Jascha Karp, Markus Kättner

Aufgabe 1

a)

Ein Evaluationskorpus könnte aus Testkollektionen bestehen. Die Kollektionen wiederum enthalten jeweils

- Dokumente
- Anfragen
- Relevanzbewertungen

Um nicht für jede zu testende Anfrage auf der gesamten Dokumentenkollektion suchen zu müssen, kann Pooling verwendet werden. Die Relevanzbewertungen müssen dann nur noch auf den gebildeten Pools ausgeführt werden.

b)

1. Recall ist vor Allem dann wichtig, wenn alle relevanten Dokumente gefunden werden sollen. Das ist z.B. bei explorativer Suche zum schreiben eines Reviews von Bedeutung.
2. Precision ist dann entscheidend, wenn ein Treffer mit hoher Wahrscheinlichkeit relevant sein soll. Das ist z.B. bei der Beantwortung einer Frage wichtig ("Wie heißt der Hund bei Mickey Mouse?").

c)

$$MAP_A = \frac{1}{N} \sum_{i=1}^N AP_i = \frac{1}{4}(0,2 + 0,1 + 0,2 + 0,9) = \frac{1,4}{4} \approx 0,3$$

$$GMAP_A = \sqrt[N]{\prod_{i=1}^N AP_i} = \sqrt[4]{0,2 * 0,1 * 0,2 * 0,9} \approx 0,2$$

$$MAP_B = \frac{1}{N} \sum_{i=1}^N AP_i = \frac{1}{4}(0,3 + 0,25 + 0,3 + 0,25) = \frac{1,1}{4} \approx 0,3$$

$$GMAP_B = \sqrt[N]{\prod_{i=1}^N AP_i} = \sqrt[4]{0,3 * 0,25 * 0,3 * 0,25} \approx 0,3$$

Aufgabe 2

Aufgabe 3

Aufgabe 4

Aufgabe 5

Aufgabe 6

Aufgabe 7

a)

$$RR_1 = \frac{1}{1} = 1$$

$$RR_2 = \frac{1}{2} = 0,5$$

b)

$$R_1 = R_2 = 3$$

$$BPREF_1 = \frac{1}{R_1} \sum_{d_r} (1 - \frac{N_{d_{r1}}}{R_1}) = \frac{1}{3} (1 + 1) = \frac{2}{3}$$

$$BPREF_2 = \frac{1}{R_2} \sum_{d_r} (1 - \frac{N_{d_{r2}}}{R_2}) = \frac{1}{3} (\frac{2}{3} + \frac{1}{3} + \frac{1}{3}) = \frac{4}{9}$$

Aufgabe 8

$$d_{1Hund} = \frac{(\log_2(f_{1Hund})+1)*\log_2 \frac{N}{n_k}}{\sqrt{\sum_{l=1}^t ((\log_2(f_{1l})+1)*\log_2 \frac{N}{n_l})^2}} = \frac{(\log_2(3)+1)*\log_2 \frac{100}{40}}{\sqrt{\sum_{l=1}^t ((\log_2(f_{1l})+1)*\log_2 \frac{100}{40})^2}} = \frac{3,4173}{5,7657} = 0,5927$$

$$d_{1Katze} = \frac{(\log_2(f_{1Katze})+1)*\log_2 \frac{N}{n_k}}{\sqrt{\sum_{l=1}^t ((\log_2(f_{1l})+1)*\log_2 \frac{N}{n_l})^2}} = \frac{(\log_2(2)+1)*\log_2 \frac{100}{20}}{\sqrt{\sum_{l=1}^t ((\log_2(f_{1l})+1)*\log_2 \frac{100}{20})^2}} = \frac{4,6439}{5,7657} = 0,8054$$

$$d_{1Maus} = \text{nicht in } D_1 \text{ enthalten}$$

$$Cosine(D_1, Q) = \frac{\sum_{j=1}^t d_{1j} q_j}{\sqrt{\sum_{j=1}^t d_{1j}^2} * \sqrt{\sum_{j=1}^t q_j^2}}$$

$$d_{2Hund} = \text{nicht in } D_2 \text{ enthalten}$$

$$d_{2Katze} = \frac{(\log_2(f_{2Katze})+1)*\log_2 \frac{N}{n_k}}{\sqrt{\sum_{l=1}^t ((\log_2(f_{2l})+1)*\log_2 \frac{N}{n_l})^2}} = \frac{(\log_2(1)+1)*\log_2 \frac{100}{20}}{\sqrt{\sum_{l=1}^t ((\log_2(f_{2l})+1)*\log_2 \frac{100}{20})^2}} = \frac{2,3219}{5,5662} = 0,4171$$

$$d_{2Maus} = \frac{(\log_2(f_{2Maus})+1)*\log_2 \frac{N}{n_k}}{\sqrt{\sum_{l=1}^t ((\log_2(f_{2l})+1)*\log_2 \frac{N}{n_l})^2}} = \frac{(\log_2(1)+1)*\log_2 \frac{100}{3}}{\sqrt{\sum_{l=1}^t ((\log_2(f_{2l})+1)*\log_2 \frac{100}{3})^2}} = \frac{5,0588}{5,5662} = 0,9088$$

$$Cosine(D_2, Q) = \frac{\sum_{j=1}^t d_{2j} q_j}{\sqrt{\sum_{j=1}^t d_{2j}^2} * \sqrt{\sum_{j=1}^t q_j^2}}$$

Aufgabe 9