

A5: Beispiel für das *Learning to Rank*-Verfahren

Zunächst legt man Sets von Faktoren fest. Jeder Faktor eines Sets wird mit allen Kombinationen von Faktoren der anderen Sets getestet, nicht aber mit den anderen Faktoren seines Sets.

Zum Beispiel könnte es folgende drei Sets geben:

- 1) citation_rank
- 2) title_worddelimiter_bm25, title_stemming_bm25, ...
- 3) abstract_worddelimiter_bm25, abstract_stemming_bm25, ...

Dann wählt man sich eine Kombination aus, die man noch nicht getestet hat (was am Anfang alle sind). Die erste Kombination von Feldern wäre: citation_rank, title_worddelimiter_bm25, abstract_worddelimiter_bm25; bei der zweiten würde man anstatt abstract_worddelimiter_bm25 abstract_stemming_bm25 verwenden usw.

Angenommen, der Durchschnitt für das citation_rank Feld wäre 0,0001, für title_worddelimiter_bm25 0,3 und für abstract_worddelimiter_bm25 0,9. Die Felder sind bereits richtig geordnet, also den Durchschnittswerten der Dokumente nach aufsteigend.

Nun berechnet man die „Basis“ für die Felder (bis auf das mit dem höchsten Durchschnitt), um die Potenzen von 3 zu bestimmen, die man als Gewichtungen ausprobiert. Die Basis für den citation_rank wäre 8, weil $\log_3 \frac{0.9}{0.0001}$ gerundet 8 ist. Die Gewichtungen für diesen Faktor wären also 3^3 , 3^4 ... 3^{13} , die für den title_worddelimiter_bm25 wären 3^{-4} , ... 3^6 .

Man berechnet nun für die erste Kombination von Gewichtungen (citation_rank: 3^3 , title: 3^{-4} , abstract: 1) die Scores für das Retrievalmaß, das man zur Evaluation verwendet. Dann wiederholt man das für die nächste Kombination von Gewichtungen (citation_rank: 3^3 , title: 3^{-3} , abstract: 1) und für alle anderen Kombinationen und notiert die Kombination mit dem höchsten Score.

Angenommen, die optimale Gewichtung lautet: 3^3 für den citation_rank und 1 (3^0) für den title. Dann würde man noch den Score für alle Kombinationen der Gewichtungen $3^{2.2}$... $3^{3.8}$ für den citation_rank und $3^{-0.8}$... $3^{0.8}$ für den title berechnen. Mit der Kombination mit dem höchsten Score berechnet man dann den Score für die Topics, die nicht zum Training verwendet wurden und notiert diesen.

