

Aufgabenblatt 8. Abgabe 06.07. 2011 8 Uhr, Besprechung 06.07. 2011 12 Uhr

Relevance Feedback

Aufgabe 31

Wie funktioniert Relevance Feedback? Überlegen Sie sich, in welchen Fällen das Relevance-Feedback-Verfahren nicht sinnvoll ist?

Aufgabe 32

Gegeben seien die folgenden Dokumente mit gewichteten Termen:

Doc1: {ROT:0.4, GRÜN:0.6, BLAU:0.3, GELB:0.5, SCHWARZ:0.7}

Doc2: {ORANGE:0.6, LILA:0.8, ROSA:0.7, BEIGE:0.6, BLAU:0.4, GELB:0.2}

Doc3: {BLAU:0.2, PINK:0.6, GRAU:0.3, BEIGE:0.5, SCHWARZ:0.4}

Doc4: {ROT:0.4, LILA:0.2, WEISS:0.7, SCHWARZ:0.4}

Doc5: {GELB:0.4, LILA:0.5, GRAU:0.2, ORANGE:0.7, SCHWARZ:0.6}

Doc6: {BLAU:0.6, GRÜN:0.4}

In dieser Menge werde mit der Anfrage

Gelb:1.0, Lila:0.2

nach dem Vektorraummodell gesucht. Dabei wird als Ähnlichkeitsmaß das Skalarprodukt verwendet. Relevant markiert seien die Dokumente 2, 3 und 4.

Berechnen sie das Relevance Feedback mit der Methode von Rocchio mit Parametern $\alpha=0.7$, $\beta=0.2$ und $\gamma=0.1$

Pseudo Relevance Feedback

Aufgabe 33

- 1) Beschreiben Sie das Prinzip der automatischen Anfrageerweiterung (Pseudo Relevance Feedback), wie sie von vielen automatischen Systemen (unter anderen SMART) in TREC benutzt wurde.
- 2) Welche Annahmen liegen diesem Anfrageerweiterungsverfahren zugrunde?
- 3) Wie unterscheiden sich diese Pseudo Relevance Feedback Verfahren vom Rocchio Verfahren?

Query Expansion berechnen

Aufgabe 34

In der Aufgabe 32 war Relevance Feedback nach dem Rocchio Verfahren berechnet worden. SMART verwendet in TREC modifizierte Relevance Feedback Verfahren zur Query Expansion.

- Nutzen Sie die drei bestplazierten Dokumente zur Expansion und expandieren Sie die Query um 3 Terme, wobei die Terme nach der Summe ihrer Gewichte in den bestplazierten Dokumenten ausgewählt werden. Verwenden Sie ansonsten die Ähnlichkeitsmaße und Gewichte aus Aufgabe 32
- Wie können dabei Gewichte für die neuen Queryterme bestimmt werden?

Quelle: <http://information-retrieval.de/ifs/ubung/inhalt.html>