

Aufgabenblatt 11: TREC Systeme

Aufgabe 1 Pseudo Relevance Feedback

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

Aufgabe 2 SMART in TREC

- Beschreiben Sie das ITL Maß, das bei SMART im zweiten TREC 3 Lauf verwendet wurde. Erklären Sie dabei die Funktionen der einzelnen Teile des Ähnlichkeitsmaßes.
- Geben Sie die einzelnen Schritte zur Berechnung der Ähnlichkeiten zwischen einer Anfrage und den Dokumenten der Sammlung an.
- Beschreiben Sie die Probleme, die sich aus der Verwendung des Cosinusmaßes für SMART in TREC 3 ergaben.
- Beschreiben Sie das Verfahren mit dem für die Dokumente aus TREC 3 der Zusammenhang zwischen der Länge, der Relevanzbeurteilung bezüglich einer Query und der berechneten Ähnlichkeit zur Query für SMART untersucht wurde.
- Welche Folgerungen können aus den Beobachtungen gezogen werden?
- Wie wurde die Gewichtsformel aufgrund dieser Beobachtungen verändert?

Aufgabe 3 Anfrageerweiterung mit Termpaaren und Teildokumenten

- Wie wurden Termpaare in TREC zur Erweiterung von Anfragen verwendet?
- Welche Ergebnisse wurden dabei erzielt?
- Wie wurden Teile von Dokumenten in TREC zum Pseudo Relevance Feedback und zur Bestimmung von Ähnlichkeiten zwischen Anfragen und Dokumenten verwendet?
- Welche Ergebnisse wurden dabei erzielt?

Aufgabe 4 Query Expansion berechnen

In Aufgabe 4 Blatt 5 war Relevance Feedback nach dem Rocchio Verfahren berechnet worden. SMART verwendet in TREC modifizierte Relevance Feedback Verfahren zur Query Expansion.

- Berechnen Sie mit den Dokumenten und der Anfrage aus Aufgabe 4 Blatt 5 die Query Expansion nach dem modifizierten Verfahren und geben Sie die Ergebnisliste an. Verwenden Sie dabei keine Termpaare. Nutzen Sie die drei bestplatzierten Dokumente zur Expansion und expandieren Sie die Query um 3 Terme, wobei die Terme nach der Summe ihrer Gewichte in den bestplatzierten Dokumenten ausgewählt werden. Verwenden Sie ansonsten die Ähnlichkeitsmaße und Gewichte aus Aufgabe 4 Blatt 5.
- Wie können dabei Gewichte für die neuen Queryterme bestimmt werden?

Aufgabe 5 Spreading Activation

- a) Beschreiben Sie das Spreading Activation Netz, das dem TREC 3 Beitrag von PIRCS zugrunde liegt.
- b) Beschreiben Sie die Ähnlichkeitsformel, die für PIRCS in TREC 4 verwendet wurde.
- c) Beschreiben Sie den Zusammenhang zwischen den beiden Darstellungsweisen.

Hinweis

Die Fachschaft Informatik führt auch in diesem Semester wieder eine Fragebogenaktion zur Evaluierung der Veranstaltungen durch; diesmal allerdings nur im WWW. Bitte schauen Sie sich die entsprechende WWW Seite an: <http://www.fachschaft.informatik.tu-darmstadt.de/feedbackneu>