

**Einführung in IR - Winter 2006/07**

Dipl.-Inform. Sascha Kriewel, LF 137

[sascha.kriewel@uni-due.de](mailto:sascha.kriewel@uni-due.de)

**Übungsblatt IR - Heimaufgabe**

10.1.2007

---

Durch die Bearbeitung und Abgabe dieser Hausarbeit können bis zu 15 Bonuspunkte für die Klausur zur Vorlesung erworben werden. Dabei sind einige Dinge zu beachten.

Die Bearbeitung der Hausaufgabe kann in Gruppen von maximal zwei Studierenden erfolgen und ist dann gemeinsam abzugeben. Identische Abgaben oder Abgaben von mehr als zwei Personen werden nicht berücksichtigt. Bei der Abgabe sollten Namen, Matrikelnummern und E-Mail-Adressen deutlich und an leicht zu findender Stelle angegeben werden.

Die Abgabe kann entweder per E-Mail an [sascha.kriewel@uni-due.de](mailto:sascha.kriewel@uni-due.de) (bevorzugt) oder im Briefkasten vor dem LF140 geschehen. Bei der Abgabe per Mail bitte ich um Dateien als PDF, bei der Abgabe auf Papier bitte nach Möglichkeit keine handschriftlichen Einreichungen (Ausnahme: Diagramme).

Für Fragen zu den Aufgaben sollte die Übungsstunde am 18. Dezember oder der Mailverteiler [ir\\_ws06@is.informatik.uni-duisburg.de](mailto:ir_ws06@is.informatik.uni-duisburg.de).

**Aufgabe 1: Vorbereitung**

Für die Bearbeitung dieses Übungsblatts sollte eine der folgenden Aufgaben ausgewählt werden:

- (a) Kontroverse zwischen Kreationisten und wissenschaftlichen Vertretern der Evolutionslehre
- (b) Kontroverse über die Eignung von Java als Lehrsprache an Universitäten oder Schulen

Nimm an, dass Du ein Referat oder eine Semesterarbeit zu diesem Thema anfertigen musst. Daher benötigst Du Dokumente zu verschiedenen Blickwinkeln und Aspekten der Suchaufgabe, Pro- und Kontraargumente, sowie Informationen aus vielfältigen Quellen (Zeitungsartikel, Blogs, Webseiten, wissenschaftliche Artikel, Diskussionsgruppen, Newsgroups, etc.).

Bereite die Suche zunächst vor, ohne eine Suchanfrage zu stellen, und halte Deine Überlegungen dabei schriftlich fest:

- (a) Arbeite die einzelnen Konzepte bzw. Aspekte der Suchaufgabe heraus.
- (b) Sammle zu den einzelnen Aspekten geeignete Terme. Zeige auf, wie Du einen Thesaurus (z.B. <http://thesaurus.reference.com/>) nutzen kannst, um Deine Suche zu verbessern.

- (c) Wähle geeignete Quellen aus, mit denen Du Deine Suche durchführen möchtest (Suchmaschinen, Zeitungsarchive, ...). Beschreibe die von Dir gewählten Quellen kurz bezüglich der Art und des Aufbaus der suchbaren Informationsobjekte und der angebotenen Suchmöglichkeiten.

### **Aufgabe 2: Suche**

Führe nun die eigentliche Suche(n) durch. Beschreibe dabei Dein Vorgehen und gib die Suchanfragen an, welche Du benutzt, sowie URLs der von Dir als relevant beurteilten Ergebnisse. Dabei wirst Du wahrscheinlich verschiedene Kombinationen der gewählten Terme versuchen und Deine ursprünglichen Terme anpassen oder ergänzen müssen.

Begründe, warum Du die Suchterme auf welche Art verknüpfst, und erlautere, welche Verbesserung Du durch (iterative) Änderungen Deiner Suchanfrage erwartest.

### **Aufgabe 3: Evaluation**

Zeige beispielhaft an den Ergebnissen einer Deiner Anfragen, wie man die Qualität des Retrievalergebnisses bewerten kann. Verwende die Gesamtmenge der von Dir als relevant beurteilten Ergebnisse und setze die betrachtete Resultatliste damit an den einzelnen Rängen in Bezug. Zeichne hierzu ein geeignetes Diagramm.

### **Aufgabe 4: Anwendung von Retrievalmodellen**

Zeige beispielhaft an Ergebnisdokumenten einer Deiner Anfragen, wie das Vektorraum-Modell funktioniert. Nimm dazu die ersten zehn Dokumente des Ergebnisses als Dokumentenkollektion und bestimme die Dokumentlängen  $l_m$ ,  $al$ , die Vorkommenshäufigkeiten  $t_{f_{mi}}$  für die Terme aus Deiner Anfrage und berechne damit die Indexierungsgewichte.<sup>1</sup>

Ermittle nun zu Deiner Anfrage ein Ranking der Dokumente. Warum unterscheidet sich dieses (wahrscheinlich) von dem Ranking der Suchmaschine?

---

<sup>1</sup>Am einfachsten lassen sich die Anzahl der Worte im Dokument und die Termhäufigkeiten bestimmen, indem man den sichtbaren Text in ein neues Dokument kopiert, und dann zum Beispiel `CountWords.jar` von der Vorlesungsseite nutzt.