Fachrichtung: Informatik Stufe: Master
Dozent: Nürnberger 10.02.2005

Konzepte des Information Retrieval

Dauer: 20 Minuten

Vorbereitungsphase

Wie lange sollte man sich auf die Klausur vorbereiten?
4 Tage

Wie hast du dich vorbereitet (allein, in der Gruppe)?
teilweise zu zweit

Wurde sich mit dem Prüfer über die Themengebiete abgesprochen?
auf XML wird nicht eingegangen

Welche Literatur/Skripte waren hilfreich?
Vorlesungsskripte reichen

Wo lagen Deiner Meinung nach besondere Schwierigkeiten der Klausur?
Was ist der Unterschied zwischen IR und Datenbanksystem?
Funktionsweise Propabilistisches Modell

Verlauf der Prüfung

Wie verlief die Prüfung?
Die verschiedenen Konzepte wurden abgefragt.
Prof. Nürnberger erwartet für eine 1, dass man alles von sich aus erzählt (Überblick geben über vorgestellte Konzepte).

Wie reagierte der Prüfer, wenn Fragen nicht sofort beantwortet wurden?
Formuliert Frage um oder gibt Lösung.

Dein Kommentar zur Prüfung:
Angenehm

Dein Kommentar zur Benotung:
Gerecht

Welche Fragen wurden konkret gestellt?
- Was versteht man unter IR?
- Wie ist ein IR-System aufgebaut? Funktionsweise?
- Worte unterliegen bestimmter Verteilung. Wie nennt man die? Was beschreibt diese?
Zipf-Verteilung: constant = termfrequency * rank (3 Bereiche)
Anwendung im IR Prozess?
Stoppworte (werden vorm Stemming gefiltert)
- Was ist der Unterschied zwischen IR und Datenbanksystem (bezogen auf Anfragen)?
- Wie werde Dokumente im IR dargestellt? Welche Modelle werden im IR angewandt? (Idee und Funktionsweisen Erklären)
Boolsches Modell
Erweitertes Boolsches Modell
Vektor-Raummodell

- Ähnlichkeitsmaße?
Euklidischer Abstand
Skalarprodukt
- Warum Normierung?
Lange Vektoren haben zu großen Einfluss, der dadurch gemindert wird
- Gibt es Nachteile bei einen der Verfahren?
Ja, Euklidischer Abstand betrachtet Vorhandensein und fehlen von Werten gleich
- Weitere Modelle?
TF-IDF Modell
Fuzzy-Modell
Propabilistisches Modell
- Was versteht man unter Clustering? Wo wird es bei IR eingesetzt?
- Welche Clustering-Verfahren gibt es? Erkläre Funktionsweise!
- Was sind moderne Techniken des Multimedia IR? (bezieht sich auf MPEG7)
- Was kann aus Audio bzw. Video extrahiert werden. Was kann annotiert werden?
- Was versteht man unter Content-based IR?
Kurze Erklärung: (da dies aus dem Skript nicht so deutlich wird)
normales IR: eingegebener Anfrage (z.B. Terme) wird Dokumentenindex (enthält auch Terme) verglichen.
content-based: eingegebene Anfrage muss in gleiche Form wie Index gebracht werden, bevor verglichen werden kann.
Bsp.:
Index einer Bildersammlung enthält Farbinformationen.
Aus der Anfrage (in Form eines Beispielbildes) müssen zum Vergleich erst die Farbinformationen ermittelt werden.
Termbasierte Anfragen auf Termindex ist auch content-based IR. Es wird aber hauptsächlich beim Multimedia Retrieval dieser Unterschied betont.