

Einführung in Information Retrieval

Prof. Dr.-Ing. Norbert Fuhr

Sascha Kriewel
sascha.kriewel@uni-due.de

23. Oktober 2006

Zielgruppen

- **Kommedia Bachelor:** Teil der Veranstaltung „Datenbanken, Informationssystem und Information Retrieval“
- **DAI Hauptstudium:** Veranstaltung im Bereich D für 6 Kreditpunkte

Leistungsnachweis

- Leistungsnachweis in Form einer Klausur
- Ende des Semesters, 90 Minuten
- für Kommedia: Note als Mittelwert aus zwei Klausuren

Übungen

- Übungszettel jede Woche (ab der nächsten)
- online als PDF; dienstags oder mittwochs
- keine Abgabe
- in der Übung: Vorstellung von Lösungen und Besprechung

Bonuspunkte

- Hausarbeit in zweiter Semesterhälfte
- alternativ: Mitarbeit bei einer Evaluation
- Bearbeitung in Zweiergruppen möglich
- bis zu 15 Bonuspunkte (entspricht etwa einer Aufgabe)

„**Information Retrieval** (IR) deals with the representation, storage, organization of, and access to information items. The representation and organization of the information items should provide the user with easy access to the information in which he is interested.“ (Baeza-Yates & Ribeiro-Neto 1999)

„Im **Information Retrieval** (IR) werden Informationssysteme in bezug auf ihre Rolle im Prozess des Wissenstransfers vom menschlichen Wissensproduzenten zum Informations-Nachfragenden betrachtet. (...) *Vage Anfragen* sind dadurch gekennzeichnet, dass die Antwort a priori nicht eindeutig definiert ist. (...) Die Darstellungsform des in einem IR-System gespeicherten Wissens ist im Prinzip nicht beschränkt (...). Die *Unsicherheit (oder die Unvollständigkeit) des Wissens* resultiert meist aus der begrenzten Repräsentation von dessen Semantik (...) Aus dieser Problematik ergibt sich die Notwendigkeit zur Bewertung der Qualität der Antworten eines Informationssystems (...)“
(Gesellschaft für Informatik, Fachgruppe „Information Retrieval“)

Anfrageverarbeitung

Bei der Anfrageverarbeitung stellt der Informations-Nachfragende eine nach festen Vorgaben eine Anfrage an das System, das die Anfrage an eine Datenbank weiterleitet. Die gefundenen Ergebnisse werden daraufhin dem Benutzer angezeigt.

IR-System?

Anfragebearbeitung

Bei der Anfragebearbeitung formuliert der Informations-Nachfragende seinen Informationswunsch als Anfrage an ein System. Das System versucht nun zu dieser Anfrage ohne weitere manuelle Eingriffe relevante Informationsobjekte zu finden.

IR-System?

Surfing

Ausgangspunkt ist hier eine Kollektion von Hypertextdokumenten, die untereinander verlinkt sind. Man startet also an einem thematisch eher allgemein ausgerichteten Knoten und bewegt sich von dort Schritt für Schritt weiter zu spezielleren Inhalten. Die Dokumente werden dabei auf einem Webserver zur Verfügung gestellt.

IR-System?

Browsing

Ausgangspunkt ist hier eine Kollektion von Hypertextdokumenten, die untereinander verlinkt sind. Man startet also an einem thematisch eher allgemein ausgerichteten Knoten und bewegt sich von dort Schritt für Schritt weiter zu spezielleren Inhalten. Die Dokumente werden dabei mit so genannten Clusterverfahren zu inhaltsähnlichen Haufen (Clustern) gruppiert.

IR-System?

Information Filtering

Mittels des Information Filtering lassen sich aus einem andauernden Strom von Dokumenten, wie bspw. Nachrichten, automatisch die für einen Benutzer interessanten herauszufiltern. Dabei wird als Grundlage meist ein sog. Benutzerprofil eingesetzt, das die Prioritäten und Informationswünsche des Benutzers speichert. Durch den kontinuierlichen Vergleich von ankommenden Dokumenten mit dem Profil werden so die relevanten Dokumente ermittelt.

IR-System?

Information Retrieval im Alltag

ist IR tatsächlich im Alltag allgegenwärtig?

- **Aufgabe:** Sammelt Gegebenheiten aus Eurem Alltag, bei denen Ihr glaubt, IR-Techniken oder IR-Systeme zu benutzen.
- jeden Tag 10 Minuten
- **Aufgabe:** Suche eines Volltextes (z.B. Daniel Rose „Information Rendezvous“) über Subito