

Allgemeine Bemerkungen zum Übungsbetrieb:

Die Übungen zur Vorlesung **Information Retrieval** im Sommersemester 2006 beginnen am 21. April und finden freitags von 12:15 bis 13:45 im LE 104 statt. Die Aufgabenblätter werden jeweils donnerstags über die Homepage der Vorlesung als PDF zum Download zur Verfügung gestellt. Zu den Übungsterminen werden die Aufgaben der Vorwoche besprochen und in Präsenzübungen einzelne Aspekte der Vorlesung vertieft. Es wird erwartet, dass die Übungsteilnehmer sich mit den Aufgaben vertraut gemacht und zumindest einen eigenen Lösungsversuch unternommen haben.

Einige Aufgaben auf den Übungsblättern sind zur Abgabe gekennzeichnet. Diese können bis zum angegebenen Termin per E-Mail abgegeben werden. Die Bearbeitung der Aufgaben in Gruppen von **2–4** Studierenden ist dabei ausdrücklich erlaubt. Bei der Abgabe (nur eine Abgabe pro Gruppe!) sind dann Namen und Matrikelnummern aller Gruppenmitglieder anzugeben. Für insgesamt 50% der möglichen Punkte aus abzugebenden Aufgaben kann ein unbenoteter Übungsschein für die Veranstaltung erteilt werden. Außerdem werden die in der Übung erreichten Punkte als Bonuspunkte in Prüfungen zur Vorlesung berücksichtigt.

Nicht zur Abgabe gekennzeichnete Aufgaben werden nicht korrigiert, sondern nur in der Übung besprochen. Zum Sammeln von Lösungen, Tipps und Vorlesungsnotizen steht ein **Wiki** zur Verfügung. Die Zugangsdaten werden in der ersten Übung bekanntgegeben. Darüber hinaus besteht die Möglichkeit, über die Mailingliste zur Übung mit anderen Teilnehmer über die Aufgaben zu diskutieren und Nachfragen an den Übungsleiter zu stellen. Den Link zum Eintragen findet man auf der Homepage der Vorlesung.

Information Retrieval - Sommer 2006

Dipl.-Inform. Sascha Kriewel, LF 137
kriewel@is.informatik.uni-duisburg.de

Übungsblatt 1

Aufgabe 1: Wissen zur Entscheidungsunterstützung

Die Fußball-Weltmeisterschaft rückt näher. Nimm an, Du möchtest Dir anlässlich der WM ein neues Fernsehgerät anschaffen. Einen Flachbildschirm soll es haben, doch darüber hinaus hast Du nur ein vages Bild von den nötigen Anforderungen. Daher soll nun eine Internetsuche weiterhelfen, damit Du Dir das für eine Kaufentscheidung nötige Wissen aneignen kannst.

Führe selbst eine Recherche durch, mit dem Ziel, die Anforderungen zu konkretisieren und letztendlich eine Entscheidung für ein bestimmtes Gerät zu treffen. Wie sind die in der Vorlesung genannten Qualitätskriterien bezüglich dieser Aufgabe zu beurteilen? (Was ist Verfügbarkeit? Was ist Inpretierbarkeit? ...)

Aufgabe 2: Anwendungen für Information Retrieval
Abgabe bis 20. April 2006, 12 Uhr:
iruebg-abgaben@is.informatik.uni-duisburg.de

In der Vorlesungen wurden einige IR-Anwendungen genannt:

- Internet-Suchmaschinen
- Internet-Kataloge
- Digitale Bibliotheken
- Suche in Nachrichten-, Mail-, News-Archiven
- Suche in Online-Dokumentationen
- Suche in Bildarchiven

Finde für jede der genannten Anwendungen mindestens drei *eigene* Beispiele im WWW (z.B. Google als Beispiel für eine Web-Suchmaschine). Wähle *jeweils* eines der Beispiele aus und untersuche es genauer.

- (a) Beschreibe, wie ein *Dokument* in dem jeweiligen System aussieht. Was für Eigenschaften besitzt es?
- (b) Bezüglich welcher dieser Eigenschaften können *Anfragen* formuliert werden?
- (c) Welche der drei *Sichten* auf Dokumente (logische Sicht, Layout-Sicht und Inhalts-Sicht) sind im System repräsentiert?

20 Punkte

Aufgabe 3: Precision vs. Recall

Precision und Recall sind zwei wichtige Retrievalmaße. Nicht immer ist es möglich, Precision und Recall gleichermaßen zu verbessern. Überlege Dir Einsatzbereiche für Text- oder Bildretrieval, in denen Du eher eine hohe Precision oder eher einen hohen Recall bevorzugen würdest. Wie sieht es bei den folgenden Beispielen aus?

- Für eine private Präsentation stellst Du Informationen zu Urlaubsorten zusammen und möchtest jeden Ort mit ein oder zwei passenden Fotos illustrieren. Dazu suchst Du in einem Fotoarchiv.
- Ein Arzt sucht in einer Digitalen Bibliothek für medizinische Texte nach passenden Diagnosen für die Symptome seines Patienten.
- Nach einer Namensänderung möchte jemand möglichst alle Webseiten finden, die noch den alten Namen benutzen, um die Betreiber anzuschreiben und auf die Änderung hinzuweisen.
- Eine Rechtsanwältin hat nur wenig Zeit vor einem Termin und möchte vorher noch möglichst viele verwandte Fälle lesen. Zur Suche benutzt sie ein Information-Retrieval-System auf juristischen Volltexten. Ob ein Fall tatsächlich relevant ist, stellt sich zumeist erst beim eigentlichen Lesen heraus.