

Information Retrieval
Prof. Dr. N. Fuhr

Sascha Kriewel
sascha.kriewel@uni-duisburg.de

Abgabe bis 16. Juni 2004

Übungsblatt Nr. 6

Aufgabe 1: Probabilistisches Retrieval - BIR-Modell

Gegeben seien die folgenden simplifizierten Dokumente mit den Termen $a, b, c, d, e, f, g, h, i, j, k, l$:

$$\begin{array}{ll}
 d_1 = a d i & d_6 = a d h l \\
 d_2 = a b i k & d_7 = a c e f h j \\
 d_3 = a d f i & d_8 = a d e h \\
 d_4 = a b c d e f g i l & d_9 = a b c e f g j k l \\
 d_5 = a b i j k l & d_{10} = a c h k
 \end{array}$$

Zu den Anfragen $q_1 = (e, f, g)$, $q_2 = (d, e, f)$, $q_3 = (b, e, f, g)$, und $q_4 = (a, c, i)$ gibt der Benutzer folgende Relevanzbeurteilungen ab:

d_i	1	2	3	4	5	6	7	8	9	10
$r(q_1, d_i)$	\bar{R}	\bar{R}	\bar{R}	R	\bar{R}	\bar{R}	R	R	R	\bar{R}
$r(q_2, d_i)$	\bar{R}	R	\bar{R}	R	\bar{R}	\bar{R}	R	R	\bar{R}	\bar{R}
$r(q_3, d_i)$	\bar{R}	\bar{R}	\bar{R}	R	\bar{R}	R	R	R	R	R
$r(q_4, d_i)$	\bar{R}	\bar{R}	R	R	R	R	R	\bar{R}	\bar{R}	\bar{R}

- Berechne die Termgewichte c_{ik} .
- In welcher Reihenfolge werden die Dokumente auf Grundlage dieser Werte gerankt?
- Wie ändert sich die Rangfolge, wenn man nur die Bewertungen der fünf besten in (b) ermittelten Dokumente berücksichtigt?
- Berechne die Wahrscheinlichkeiten $P(R|q, \vec{x})$ für die Dokumente auf den beiden möglichen Wegen (direkt/über das BIR-Modell) und vergleiche die Ergebnisse. Wodurch ist der Unterschied zu erklären?

15 Punkte

Aufgabe 2: Interfaces und Visualisierung

Bearbeite **eine** der beiden folgenden Alternativen.

Alternative 1: Benutzerinterface für die gesamte Suche

Bei der Entwicklung des DAFFODIL-Systems (Zugriff unter <http://www.daffodil.de>) wurde versucht, sich an bestehenden Interfacedesign-Modellen zu orientieren. Experimentiere mit dem System und beurteile inwieweit es den Prinzipien des WOB-Modells folgt. Begründe Deine Meinung jeweils kurz (zwei bis drei Sätze reichen).

Welchen anderen, in der Vorlesung besprochenen Designprinzipien für ein Benutzerinterface zur Suche folgt das System, bzw. gegen welche verstößt es?

15 Punkte

Alternative 2: Visualisierung in Informationssystemen

Ben Shneiderman postuliert in seinem bekannten Artikel zur Informations-Visualisierung „The Eyes Have It“ das sogenannte *Visual Information-Seeking Mantra*:

„Overview first, zoom and filter, then details-on-demand.“

Laut Shneiderman ist dies das grundlegende Prinzip bei der Erstellung von Visualisierungssystemen für die Informationssuche. Wähle Dir ein bestehendes System zur Informations-Visualisierung. Nimm dabei die Systeme aus der Vorlesung als Anhaltspunkt zur Suche, wenn Du selbst kein Visualisierungssystem kennst.

Versuche Shneidermans *Visual Information-Seeking Mantra* darauf anzuwenden. Folgt das System diesem Prinzip? Wo weicht es davon ab? Setze Dich auch kritisch mit dem Mantra auseinander.

Der Volltext des Artikels steht unter

[ftp://ftp.cs.umd.edu/pub/hcil/Reports-Abstracts-Bibliography/
96-13html/96-13.html](ftp://ftp.cs.umd.edu/pub/hcil/Reports-Abstracts-Bibliography/96-13html/96-13.html)

Hier ist noch zwei Übersichtsseiten mit Ressourcen zur Informations-Visualisierung im WWW (leider beide etwas veraltet):

[http://graphics.stanford.edu/courses/cs348c-96-fall/resources.
html](http://graphics.stanford.edu/courses/cs348c-96-fall/resources.html)

<http://www.otal.umd.edu/Olive/>

15 Punkte