

**Klausur zur Vorlesung
Information Retrieval
(WS 2006 / 2007, LV-Nr. 36 600)**



**im Studiengang Informationswissenschaft
Montag, 12. Februar 2007**

Prof. Dr. Christian Wolff
Professur für Medieninformatik
Institut für Medien-, Informations- und Kulturwissenschaft

Allgemeine Hinweise

1. Bearbeitungszeit: 90 Minuten.
2. Maximal erreichbare Punktzahl: 90. Zu Ihrer Orientierung sind die erreichbaren Punkte bei jeder Frage genannt – bitte teilen Sie die Arbeitszeit entsprechend ein.
3. Schreiben Sie Ihren **Namen, Vornamen und Ihre Matrikelnummer (oder eine frei wählbare ID)** leserlich auf alle Klausurbögen, die Sie für Ihre Lösung verwenden - **bevor** Sie mit der Bearbeitung beginnen! Blätter ohne diese Angaben können nicht gewertet werden.
4. Verwenden Sie nur die bereitgestellten Klausurbögen.
5. Haken Sie ggf. nach Bearbeitung die Aufgaben auf der Angabe ab, um sicherzustellen, dass Sie keine Frage ausgelassen haben.
6. Benutzen Sie **keine Bleistifte, keine rotschreibenden Stifte und kein TippEx** (oder ähnliche Produkte).
7. Es sind **keine** weiteren Unterlagen (Skripte, Vorlesungsmitschriften, etc.) zugelassen.
8. Wenden Sie sich bei Unklarheiten in den Aufgabenstellungen immer an die Aufsichtsführenden. Hinweise und Hilfestellungen werden dann, falls erforderlich, offiziell für den gesamten Hörsaal durchgegeben. Aussagen unter „vier Augen“ sind ohne Gewähr.
9. Geben Sie keine **mehrdeutigen** (oder **mehrere**) Lösungen an. In solchen Fällen wird stets die Lösung mit der geringeren Punktzahl gewertet. Eine richtige und eine falsche Lösung zu einer Aufgabe ergeben also null Punkte.
10. Formulieren Sie Ihre Antworten (ggf. knapp) aus; die bloße Nennung von Stichwörtern kann als Antwort nicht gewertet werden!
11. Verändern Sie die Aufgabenstellung nicht, um Sie an Ihre Lösung **„anzupassen“**. Lösungen, die sich nicht an die vorgegebenen Aufgabenstellungen halten, werden mit null Punkten bewertet.

Fragen	Punkte
1. Erklären Sie anhand konkreter Beispiele, was man unter Information Retrieval versteht. Welche Gründe lassen sich für die wachsende Bedeutung des Information Retrieval nennen?	9
2. Vergleichen Sie das Boolesche Retrievalmodell mit dem Vektorraummodell. Erläutern Sie die Prinzipien beider Modelle und stellen Sie anschließend Gemeinsamkeiten und Unterschiede heraus.	20
3. In einem Retrievalexperiment mit einer Internetsuchmaschine werden bei einem <i>cut off</i> -Wert von 30 ein <i>recall</i> von 0,3 und eine <i>precision</i> von 0,5 für eine bestimmte Treffermenge berechnet. <ul style="list-style-type: none"> • Wie viele relevante Dokumente gibt es in der Dokumentkollektion bzw. in der Treffermenge des Benutzers und wie bestimmen Sie dies? • Erläutern Sie die Begriffe <i>cut off</i>-Wert, <i>recall</i> und <i>precision</i>! • Wie kann allgemein der <i>recall</i> in solch einem Experiment gemessen werden? 	14
4. Erklären Sie die TF-IDF-Termgewichtung. Auf welchen Annahmen beruht Sie? Erläutern Sie diese Annahmen an einem konkreten Beispiel!	13
5. Der folgende Text soll indexiert werden: Meine Suchmaschine versteht mich Das Startup-Unternehmen Powerset will eine Suchmaschine entwickeln, die in natürlicher Sprache gestellte Fragen beantwortet. Das Unternehmen hat Technologie vom Palo Alto Research Center (PARC) lizenziert. Das Forschungszentrum ist eine Tochtergesellschaft von Xerox und arbeitet seit 30 Jahren daran, Computer die Bedeutung von Texten erkennen zu lassen. Powerset hatte im Herbst 2006 12,5 Millionen US-Dollar Risikokapital von Investoren erhalten. Der PARC-Forscher Ronald Kaplan wechselt als Chief Technology zu Powerset. PARC erhält Lizenzgebühren und eine Beteiligung am Startup-Unternehmen. <small>Quelle: Heise-Verlag (Hrsg.) (2007). Heise Online News. Meine Suchmaschine versteht mich, http://www.heise.de/newsticker/meldung/85107, Zugriff 02 / 07.</small> Nach welchen Kriterien wählen Sie Indexterme aus bzw. gewichten Sie sie? Welche sprachlichen Phänomene sind ggf. relevant?	12
6. Definieren und erläutern Sie das Cosinus-Maß als Retrievalfunktion im statistischen Retrievalmodell. Geben Sie ein Berechnungsbeispiel für einen Anfrage- und zwei Dokumentvektoren mit fünf (Term-) Dimensionen.	10
7. Erläutern Sie die Begriffe Synonymie, Homonymie, Antonymie, Meronymie, Hyponymie und Hyperonymie jeweils anhand eines Beispiels.	12

Summe	90
--------------	-----------