

**Klausur zur Vorlesung  
Information Retrieval  
(WS 2008 / 2009, LV-Nr. 36 600)**



Universität Regensburg



**im Studiengang Informationswissenschaft  
Montag, 2. Februar 2009**

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, Studiengang (B.A. / Mag.) und Studienfach sowie 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. Beschreiben und erläutern Sie die wichtigsten Teilaufgaben im Information Retrieval und stellen sie den Dokumentationsprozess grafisch dar.	15
2. Was versteht man unter einem Retrievalmodell? Wie lässt es sich formalisieren?	10
3. Erläutern Sie anhand eines selbst gewählten Beispiels, wie im Booleschen Retrievalmodell eine Anfrage, die als Ausdruck der Booleschen Logik repräsentiert ist, abgearbeitet werden kann. Das Beispiel sollte aus einer Anfrage mit wenigstens zwei unterschiedlichen Booleschen Operatoren sowie aus einer invertierten Datei bestehen.	12
4. Erläutern Sie das Grundprinzip des Vektorraummodells und benennen Sie je zwei Vor- und Nachteile dieses Modells	12
5. Erläutern Sie den Unterschied zwischen Klassifikation und Clustering von Dokumenten.	5
6. Bei der Indexierung von Texten können Terme aus dem Dokumenttext oder aus einer externen Wissensquelle stammen. Die Zuordnung von Begriff zu Dokument kann intellektuell oder automatisch erfolgen. Geben Sie je ein Beispiel für eine der vier Kombinationsmöglichkeiten.	8
7. In einem Retrieval experiment mit einem Information Retrieval-System auf der Basis des Vektorraummodells werden bei einem cut off-Wert von 50 ein recall von 0,2 und eine precision von 0,4 für eine Treffermenge berechnet. <ul style="list-style-type: none"> <li>• Erläutern Sie die Begriffe cut off-Wert, recall und precision!</li> <li>• Wie viele relevante Dokumente gibt es in der Dokumentkollektion bzw. in der Treffermenge des Benutzers?</li> </ul>	8
8. Erläutern Sie das Evaluationsmaß der <i>Mean Average Precision</i> (Erklärung, Vorteile, Einsatzbereiche).	8
9. Wodurch unterscheiden sich Suchmaschinen für das WWW von „traditionellen“ Retrievalsystemen, wie sie z. B. im Bereich der Fachinformation zum Einsatz kommen – erläutern Sie vier wesentliche Merkmale.	12
<b>Summe</b>	<b>90</b>