

**Klausur zur Vorlesung**  
**Information Retrieval**  
**(WS 2009 / 2010, LV-Nr. 36 600)**



**im Studiengang Informationswissenschaft**  
**Mittwoch, 10. Februar 2010**

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.

<b>Fragen</b>	<b>Punkte</b>
1. Erläutern Sie an einem konkreten Beispiel die Vielfalt der mit einer informationellen Einheit potenziell verbundenen Fragestellungen und erklären Sie auf dieser Basis Besonderheiten des Information Retrieval im Unterschied zu Datenbanken.	12
2. <b>[Aus Anlass der heutigen Abschiedsvorlesung von Rainer Kuhlen]</b>  Beschreiben Sie anhand des Modells von Kuhlen Transformationsprozesse zwischen Wissen und Information und erläutern Sie an einem konkreten Beispiel, welche Rolle Information Retrieval-Systeme spielen.	12
3. Auf welchen Prinzipien baut das Boolesche Retrievalmodell auf? Erklären Sie Stärken und Schwächen an einem konkreten Beispielszenario.	14
4. Was versteht man unter <i>relevance</i> feedback und warum lässt sich durch diese Technik die Retrievaleffektivität verbessern?	8
5. Nennen und erläutern Sie zwei Kriterien, nach denen man Clustering-Algorithmen kategorisieren kann.	6
6. In einem Retrievalexperiment mit einem Information Retrieval-System auf der Basis des Vektorraummodells werden bei einem cut off-Wert von 20 ein recall von 0,1 und eine precision von 0,6 für eine Treffermenge zu einer Suchanfrage ermittelt. <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 insgesamt bzw. in der Treffermenge des Benutzers?</li> </ul>	8
7. Erläutern Sie das Grundprinzip und Arbeitsweise des <i>latent semantic indexing</i> .	10
8. Welche Faktoren beeinflussen bei der Indexierung die Auswahl und Gewichtung geeigneter Terme? Nennen Sie fünf wesentliche Faktoren.	10
9. Lohnt sich die Arbeit mit mehr als einer Suchmaschine bei der Recherche im offenen Web?  Welche Alternativen zu klassischen Suchmaschinen wie Google bieten sich an?	10
<b>Summe</b>	<b>90</b>