

Name:	Studiengang: <input type="checkbox"/> B.A. <input type="checkbox"/> MA.
Vorname:	In FlexNow angemeldet: <input type="checkbox"/> Ja <input type="checkbox"/> Nein
Matrikelnummer:	
Studienfächer:	Fachsemester Informationswissenschaft:

Allgemeine Hinweise:

1. Überprüfen Sie bitte, ob Sie alle Seiten der Klausurangabe vollständig erhalten haben (Gesamtzahl: 5)
2. **Bearbeitungszeit: 60 Minuten**, maximal erreichbare **Punktezah: 60**. Die jeweils erreichbare Punktezah ist bei jeder Frage angegeben. Bitte teilen Sie Ihre Arbeitszeit entsprechend ein.
3. Denken Sie daran, die Daten oben einzutragen, **bevor** Sie mit der Bearbeitung beginnen.
4. Treffen Sie bitte die Auswahl Ihrer Antworten bei Multiple-Choice-Fragen **direkt** auf dieser Klausurangabe.
5. Verwenden Sie für die Beantwortung der Freitext-Fragen ebenfalls diese Klausurangabe. Sie können jederzeit auch die Rückseiten beschreiben, falls der Platz auf der Vorderseite nicht ausreichen sollte. Bitte geben Sie in jedem Fall an, auf welche Frage sich die Lösung jeweils bezieht.
6. Benutzen Sie keine Bleistifte, keine rot schreibenden Stifte und kein TippEx, o.ä.
7. Keine Hilfsmittel sind zugelassen d.h. keine Foliensätze oder selbstgeschriebene Notizen.
8. Ein Taschenrechner dürfen Sie gerne benutzen.
9. Mobiltelefone sowie Computer am Arbeitsplatz - auch ausgeschaltet - sind **nicht zugelassen**.
10. 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 ergeben also null Punkte.
11. Wenden Sie sich bei Unklarheiten in den Aufgabenstellungen immer an den Aufsichtsführenden. Hinweise und Hilfestellungen werden dann, falls erforderlich, offiziell für alle Teilnehmer durchgegeben.
12. Bei falschen Antworten in Multiple-Choice Fragen werden Punkte Abgesogen.

1. A document score can be estimated as follows using a language modelling approach:

Die Bewertung eines Dokuments kann in einem LM-Verfahren mit der folgenden Gleichung geschätzt werden:

$$P(d|q) \propto P(d) \prod_{t \in q} ((1 - \lambda)P(t|M_c) + \lambda P(t|M_d))$$

a) Explain what each of the highlighted components is and how it may be estimated.

Erklären Sie was die markierten Teile sind und wie sie geschätzt werden können. (4 Punkte)

b) What happens as the smoothing parameter lambda tends towards 1?

Welcher Effekt hat der Parameter Lambda als er den Wert 1 nähert? (3 Punkte)

c) Explain two ways the document prior P(d) may be estimated in a web-search context

Erklären Sie zwei Methoden, die benutzt werden könnten um das Dokument 'Prior' zu schätzen im Zusammenhang einer Websuchmaschine. (6 Punkte)

2a) Name two particular problems associated with performing IR on Japanese documents
Nennen Sie zwei Probleme, die man lösen muss, wenn man IR auf japanische Dokumente anwendet. (4 Punkte)

b) Is a positional index useful for the following queries? Explain each answer (hint: you may want to refer to bi-word indexes in some of your answers)

Wäre ein 'Positional Index' für die folgende Anfragen hilfreich? Begründen Sie jede Antwort (Hinweis: ein Vergleich mit 'bi-word Indexes' könnte für manche Anfragen hilfreich sein)(9 Punkte)

- airport

- "Thomas Müller"

- "Benedikt Höwedes"

d) Which of the following statement are? *Welche der folgenden Aussagen stimmen zu? (8 Punkte)*
Stimmt zu *Stimmt nicht zu*

- | | |
|--|--------------------------|
| <input type="checkbox"/> The jaccard coefficient relates to a probabilistic mathematical framework
<i>Jaccard ist auf probabilistischer Basis realisiert</i> | <input type="checkbox"/> |
| <input type="checkbox"/> The performance of jaccard in IR is limited as it has no length normalisation component
<i>ein Grund, dass jaccard für IR weniger geeignet ist, ist dass die Länge der Dokumente nicht berücksichtigt wird</i> | <input type="checkbox"/> |
| <input type="checkbox"/> The performance of jaccard in IR is limited as it fails to account for the rareness of terms
<i>ein Grund, dass jaccard für IR weniger geeignet ist, ist dass selten vorkommende Wörter nicht anders gewichtet werden.</i> | <input type="checkbox"/> |
| <input type="checkbox"/> The performance of jaccard in IR is limited as it has no concept of term-frequency | <input type="checkbox"/> |

3) The table shows result lists for a system on 3 known-item search tasks R (relevant) N (not relevant) Die Tabelle zeigt die ersten 10 Treffer von einer Suchmaschine für 3 known-item Aufgaben relevant (R) sind oder nicht (N)

	Task 1	Task 2	Task 3
1	N	N	R
2	N	R	N
3	N	N	N
4	N	N	N
5	N	N	N
6	N	N	N
7	N	N	N
8	R	N	N
9	N	N	N
10	N	N	N

Platz zum Rechnen

a) Calculate MAP over the three tasks
Berechnen Sie MAP für die Drei Aufgaben (3 Punkte)

b) Calculate a MRR score over the three tasks
Berechnen Sie MRR für die Drei Aufgaben (3 Punkte)

c) Calculate P@10 over the three tasks (3 Punkte)
Berechnen Sie ein P@10 Wert für das System für die Drei Aufgaben

d) Reason about the merits of these metrics for in the context of a known-item.
Erläutern Sie Ihre Gedanken zu der Angemessenheit der Drei Metriken für known-item Retrieval. (6 Punkte)

5) Describe the main ideas behind Savolainen's ELIS model and relate these to information retrieval

Beschreiben Sie die Hauptideen hinter Savolainens ELIS-Modell ein Modell. Erklären Sie was das Modell mit Information Retrieval zu tun hat. (5 Punkte)

6) Consider a web graph with three nodes 1, 2 and 3. The links are as follows: Stellen Sie sich ein Webgraph mit 3 Knoten, 1, 2 und 3 und den folgenden Verbindungen vor:

$1 \rightarrow 2, 3 \rightarrow 2, 2 \rightarrow 1, 2 \rightarrow 3.$

a) Provide the transition probability matrix for the surfer's walk with teleporting when $\alpha = 0.5$

Geben Sie die Übergangswahrscheinlichsmatrix für den Graph, wenn Teleporting benutzt wird und $\alpha = 0.5$.
(3 Punkte)

b) If we start the random walk on node 1, what are the probability of being in nodes 1, 2 and 3 after 2 transitions?. Falls wir auf Knoten 1 anfangen, mit welcher Wahrscheinlichkeit werden wir auf Knoten 1, 2 und 3 nach 2 Schritten sein?
(3 Punkte)