

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: 90 Minuten**, maximal erreichbare **Punktzahl: 65**. Die jeweils erreichbare Punktzahl 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. Einen 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. Bei Multiple-Choice-Fragen Führen falsche Antworten zu Punktabzug
12. Wenden Sie sich bei Unklarheiten in den Aufgabenstellungen immer an den Aufsichtsführenden. Hinweise und Hilfestellungen werden dann, falls erforderlich, offiziell für alle Teilnehmer durchgegeben.

1) Welche der genannten Systeme machen die folgenden Annahmen? (10 Punkte)
 Markieren Sie alle passenden Antworten. Falsche Markierungen führen zu Punktabzug

Which of the above systems make the following assumptions? Mark all that apply:

<ul style="list-style-type: none"> Relevanzbeurteilungen sind voneinander unabhängig. <i>Relevance judgments are independent.</i> 	LM with JM smoothing	BIM	BM25	VSM
<ul style="list-style-type: none"> Wörter, die nicht in der Abfrage sind, kommen mit einer gleichen Wahrscheinlichkeit in relevanten und nicht-relevanten Dokumenten vor. <i>Terms not in the query are equally likely to appear in relevant and non-relevant documents.</i> 	LM with JM smoothing	BIM	BM25	VSM
<ul style="list-style-type: none"> die Wahrscheinlichkeiten des Auftretens zwei verschiedene Wörter in einem Dokument sind unabhängige Ereignisse. <i>The probability of terms appearing in a document are treated as independent</i> 	LM with JM smoothing	BIM	BM25	VSM
<ul style="list-style-type: none"> Relevanz ist unabhängig von Termfrequenz. <i>relevance is independent of term frequency</i> 	LM with JM smoothing	BIM	BM25	VSM
<ul style="list-style-type: none"> Die Länge eines Dokuments beeinflusst nicht die Wahrscheinlichkeit, dass ein Wort in dem Dokument vorkommt. <i>The length of a document does not influence the probability of a term being in a document</i> 	LM with JM smoothing	BIM	BM25	VSM

2) Welche der folgenden Aussagen stimmen zu? Which of the following statements are true? (10 Punkte)

Wahr		Falsch
<input type="checkbox"/>	Stemming vergrößert die Länge der Postingslisten <i>stemming increases the size of the postings lists</i>	<input type="checkbox"/>
<input type="checkbox"/>	Stemming senkt nie Recall <i>stemming never lowers recall.</i>	<input type="checkbox"/>
<input type="checkbox"/>	IDF beeinflusst die Ergebnisse von Einzelwort-Anfragen z.B. wie "apple" <i>idf can influence the scores for single term queries like "apple".</i>	<input type="checkbox"/>
<input type="checkbox"/>	Die Entfernung von Stopwörtern verkleinert den Index <i>Stopword removal reduces the size of the index</i>	<input type="checkbox"/>
<input type="checkbox"/>	Normalisierungsverfahren z.B. "case-folding" sollten nur an Dokumenten durchgeführt werden und nicht an Anfragen <i>normalisation processes, such as case-folding only need to be applied to documents and not queries</i>	<input type="checkbox"/>

3a) Wenden Sie das BIM-Modell an, um für die Suchanfrage "information class" ein Ranking für die unten aufgeführten Dokumente auszurechnen. Es gibt nur diese 5 Dokumente in der Sammlung. *Using the BIM model calculate a ranking for the following documents given the query "information class" (we assume that these are the only documents in the collection)* (10 Punkte)

D1: "information retrieval is the best class I have taken in Regensburg"

D2: "I learned in information retrieval class that long documents like this one have an advantage and are always ranked higher than they should be"

D3: "Robotics and information retrieval make use of the same technologies"

D4: "information retrieval existed long before google"

D5: "After this exam I am going for a beer. This class was tough and so was the exam".

b) Nennen Sie einen Unterschied im Ranking, den es gegeben hätte, falls BM25 anstelle von BIM benutzt worden wäre (Sie müssen nichts rechnen). (5 Punkte)

Without calculating anything, name one obvious difference in the ranking if BM25 had been used instead:

4) Erklären Sie wie einen der ihnen bekannten Information-Retrieval Algorithmen benutzt werden kann, um zu schätzen, welche Rezepte in einer Sammlung ein Benutzer schmackhaft finden würde. Stellen Sie sich vor, dass der Benutzer schon 10 Rezepte als entweder "schmackhaft" oder "nicht-schmackhaft" bewertet hat. Ihre Antwort sollte erklären, wie Rezepte und die Präferenzen eines Nutzers modelliert werden können und wie Scores für ein (Rezept-Benutzer) Paar berechnet werden können. (10 Punkte)

Explain in detail how any information retrieval model you know about can be used to estimate which recipes in a given collection a user will rate as tasty. Imagine that a user has already rated 10 recipes as either "tasty" or "not-tasty". Explain how recipes and the user's preferences could be modelled and how scores for a (recipe, user) combination could be calculated.

5) Nehmen Sie an, dass das System, das Sie in 4) beschrieben haben, als Ranking die Top-10 Rezepte für den Benutzer liefert. Erklären Sie wie man mit Hilfe von „IR-Performance-Metrics“ die Leistung des Systems evaluieren kann. Nennen Sie ein passendes Maß und begründen Sie Ihre Antwort. Erklären Sie an den unten gezeigten Ergebnissen wie man das Maß berechnen kann. (10 Points)

*Assume the system you described in 4) returns the top 10 recipes for a user as an **ordered ranking**. Explain how we might evaluate the system using IR performance metrics. Name one suitable metric and justify your choice. Describe how the metric would be calculated given the following example.*

Vorschläge für TestUser 1

1. Käsespätzle mit Feldsalat
2. Pizza mit Erdbeern
3. Rösti mit Apfelmus
4. Ente mit Orangensauce und Reis
5. Gemüseragout mit Hackbällchen
6. Pikante Reispfanne
7. Fischfilet in Senfrahm
8. Lancashire Hotpot
9. Nudelrisotto mit Gemüse
10. Möhren - Ingwer - Suppe

Vorschläge für TestUser 2

1. Pikante Reispfanne
2. Fischfilet in Senfrahm
3. Lancashire Hotpot
4. Nudelrisotto mit Gemüse
5. Möhren - Ingwer – Suppe
6. Käsespätzle mit Feldsalat
7. Pizza mit Artischocken
8. Pilz-risotto
9. Fish und Chip
10. Kraut und Schupfnudeln

6) Markieren Sie alle der unten genannten Suchbenutzeroberflächenkomponenten, die das Problem der Mehrdeutigkeit gezielt lösen. Falsche Markierungen führen zu Punktabzug (10 Points)

Which of the following SUI features aim to solve the issue of polysemy? Mark all that apply.

Clustering results (*Ergebnisse Clustern*)

Query suggestions (*empfohlene Anfragen*)

Speech as Input (*Spracheingabe*)

Faceted meta-data

Breadcrumb bar (*Breadcrumbsleiste*)