

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: 6)
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. 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.

1a) What does personalisation mean in the context of IR?

Was bedeutet Personalisierung im Zusammenhang mit Information Retrieval? (3 Punkte)

b) Describe one way in which personalisation may be implemented on an IR system.

Beschreiben Sie eine Methode, mit der Personalisierung umgesetzt werden kann. (3 Punkte)

c) Describe one potential practical problem with personalisation in IR from the point of view of the user.

Beschreiben Sie ein praktisches Problem im Zusammenhang mit Personalisierung vom Standpunkt des Nutzers aus gesehen (4 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"

c) Is it sensible to apply stemming to both queries and documents or is it sufficient and best only to stem the query? Explain your answer.

Ist es sinnvoll 'stemming' nur auf die Anfragen anzuwenden oder sollte man Wörter in den Anfragen und Dokumenten auf den Wortstamm reduzieren? Begründen Sie Ihre Antwort(4 Punkte)

3 a) Circle the retrieval models for which the following statements are true:
 Markieren Sie die Modelle, auf die die folgende Aussagen zutreffen:(8 Punkte)

- The model makes use of term frequency: (BIM) (BM25) (LM with JM Smoothing) (VSM)
 Das Modell verwendet Termfrequenz:
- The model makes use of collection frequency: (BIM) (BM25) (LM with JM Smoothing) (VSM)
 Das Modell verwendet 'collection frequency':
- The model has an IDF or IDF-like feature: (BIM) (BM25) (LM with JM Smoothing) (VSM)
 Das Modell verwendet IDF oder eine IDF-ähnliche Eigenschaft:
- The model uses length normalisation: (BIM) (BM25) (LM with JM Smoothing) (VSM).
 Das Modell erfasst Längennormalisierung:

c) Complete the tables below for the documents D1-D5 **without** using any form of normalisation
 Füllen Sie die folgenden Tabellen mit Hilfe der Dokumente D1-D5 aus. Verwenden Sie dabei **keine** Normalisierungsverfahren.(3 Punkte)

Q: reus driving fine
 D1: marco reus drives defenders crazy with his tricks (non relevant)
 D2: reus given fine for driving without license (relevant)
 D3: it is a fine feeling to drive a rolls royce (non relevant)
 D4: the weather outside is frightening (non relevant)
 D5: it is a fine time to go shopping (non relevant)

reus	Present	Not present		driving	Present	Not present		fine	Present	Not present
Relevant			Relevant				Relevant			
Non relevant			Non relevant				Non relevant			

d) If we were to use these tables to estimate term-weights for BIM would the term 'reus' receive a positive, negative or neutral weight?

Falls wir die Tabellen nutzen würden, um die Term-Gewichte zu schätzen, würde 'reus' ein positives, negatives oder neutrales Gewicht bekommen? (3 Punkte)

e) Using the same approach would the term 'fine' have a higher, lower or equal weight than 'reus'?
 Würde 'fine' mit demselben Verfahren ein höheres, niedrigeres, oder gleiches Gewicht als 'reus' erhalten? (3 Punkte)

f) Why can BIM not be estimated using the approach in c) and d) in practice? What is typically done instead?

Erklären Sie warum BIM in der Tat nicht wie in c) und d) geschätzt werden kann und was typischerweise stattdessen gemacht wird. (4 Punkte)

4a) Explain how IR researchers typically lower the costs of attaining relevance assessments for large test collections. Name one problem with the solution.

Erklären Sie wie IR-Forscher typischerweise die Kosten der Erhebung von Relevanzbeurteilungen für grosse Testsammlungen reduzieren können. *Nennen Sie ein Problem mit dem Verfahren. (4 Punkte)*

b) Describe the main concepts of Bates' Berrypicking model. In which ways does this model contradict the assumptions made in systems IR (Cranfield paradigm)?

Erklären Sie die Hauptkonzepte des Berrypicking-Modells nach Bates. Inwiefern widerspricht dieses Modell den zentrale Annahmen des System-IR (Cranfield Paradigma)? (4 Punkte)

5a) Name a search user interface feature designed to help with polysemy
Nennen Sie ein Suchoberflächenfunktion, die bei Polysemie hilft. (2 Punkte)

b) What is Rocchio's algorithm for and how does it work?
Wozu wird Rocchios Algorithmus benutzt und wie funktioniert er? (5 Punkte)

c) Name a practical (implementation) problem for Rocchio's algorithm and how we can deal with this.
Nennen Sie ein praktisches Problem, das beim Einsatz von Rocchios Algorithmus vorkommen kann. Wie kann man damit umgehen? (2 Punkte)