Name:	**Studiengang: ☐ B.A. I ☐ MA.**
Vorname:	**In FlexNow angemeldet: ☐ Ja I ☐ 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 **Punktezahl: 65**. Die jeweils erreichbare Punktezahl 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.

1a) Explain what is meant by the "feast or famine" problem in Boolean retrieval. Relate each to a query operator.
Erklären Sie das 'Feast or Famine' Problem im Zusammenhang von boolischem Retrieval. Was hat es mit bestimmten Anfrage-Operatoren zu tun? (4 Punkte)

b) Does the following merge algorithm perform an "AND", an "OR" or an "AND NOT" operation?
Führt der unten dargestellte Algorithmus eine "AND", "OR" oder "AND NOT" Operation durch?
(3 Punkte)

```
INTERSECT(p1, p2)
 1  answer ← ⟨ ⟩
 2  while p1 ≠ NIL and p2 ≠ NIL
 3  do if docID(p1) = docID(p2)
 4        then ADD(answer, docID(p1))
 5             p1 ← next(p1)
 6             p2 ← next(p2)
 7        else if docID(p1) < docID(p2)
 8                then p1 ← next(p1)
 9                else p2 ← next(p2)
10  return answer
```

c) For BIM to achieve comparable performance to BM25, which of the following are required?
Welche Bedingungen sind erforderlich damit BIM ähnliche Performanz wie BM25 leisten kann?(6 Punkte)

- large collection
eine große Sammlung
- small collection
eine kleine Sammlung
- collection size is unimportant
Größe der Sammlung ist unwichtig

- long documents
lange Dokumente
- short documents
kurze Dokumente
- document length is unimportant
Länge der Dokumente ist unwichtig

- appropriately set parameters
richtig gesetzte Parameter
- inappropriately set parameters
falsch gesetzte Parameter

Explain this last answer: *Erklären Sie Ihre Antwort auf die letzte Frage:*

d) Which properties are true for the vector space model? *Welche der folgenden Aussagen stimmen für das Vektorraummodell zu? (6 Punkte)*

- the length of the document can be accounted for in the model
das *Modell kann die Länge der Dokumente berücksichtigen*

- the frequency with which a query term features in a document affects a document score
wie oft ein Wort in einem Dokument vorkommt beeinflusst wie hoch ein Dokument bewertet wird

- rarer terms are typically weighted as less important
selten *vorkommende Wörter werden niedriger gewichtet*.

- 2) 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)

3) The table below lists relevance judgements for the top 10 documents returned by 3 systems.
Die Tabelle zeigt ob die ersten 10 Treffer von drei Suchmaschine relevant (R) sind oder nicht (N)

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

Platz zum Rechnen

a) which of the 3 systems has the ***highest*** average precision score?
*Welches System leistet die **beste** 'Average precision'? (3 Punkte)*

System 1 System 2 System 3

b) which of the following 3 systems has the ***lowest*** average precision score?
*Welches System leistet die **schlechteste** 'Average precision'? (3 Punkte)*

System 1 System 2 System 3

c) which metric best describes the comparative performance of these systems for informational web search? Explain your answer:
Welches Maß beschreibt am besten die Leistung der 3 Systeme? Erklären Sie Ihre Antwort. (4 Punkte)

P@1, P@5, P@10, Average precision

4a) Complete the following table for documents D1 and D2. There are 100,000 documents.
Füllen Sie die Tabelle für Dokumente D1 und D2 aus. Es gibt 100,000 Dokumente (7 Punkte)

Term	**DF**	**IDF**	**TF-D1**	**TF-IDF-D1**	**Euclid. Norm. D1**	**TF-D2**	**TF-IDF-D2**	**Euclid. Norm. D2**
bus	500							
ticket	200							
regenburg	50							
tourist	100							

Q	bus ticket regensburg
D1	bus bus bus ticket
D2	regensburg regensburg tourist

b) Which document ranks higher using term weights as calculated in a) and with query weighting of 1 if a term is present in the query and 0 otherwise?
Welches Dokument wird höher gerankt, falls man Wortgewichtungen wie in a) berechnet und eine Querygewichtung von 1 gibt, wenn ein Wort ein Teil der Anfrage ist und 0 wenn nicht. (5 Punkte)

5) Name and describe a model of information behaviour that describes the relation between information retrieval from a systems perspective and information seeking behaviour.
Beschreiben Sie ein Modell des Informationsverhaltens, das die Beziehung zwischen Information Retrieval aus der System-Perspektiv und Informationsuchverhalten erläutert. (4 Punkte)

6) Name 7 features that a search engine may use to estimate whether or not a document is relevant with respect to a query. (7 Punkte)
Nennen Sie 7 Merkmale, die eine Suchmaschine benutzen kann um zu schätzen ob ein Dokument für eine gegebene Suchanfrage relevant ist.