

Prof. David Elsweiler

16.07.2018

## WHK "Einführung in die Informationswissenschaft" WS 2017/2018

<i>Nachname, Vorname</i>	
<i>Abschluss (BA, MA, FKN etc.)</i>	
<i>Matrikelnummer, Semester</i>	
<i>Versuch (1/2/3)</i>	

**Bitte füllen Sie zuerst den Kopf des Angabenblattes aus!**

**In diesem Klausurteil sind insgesamt 43 Punkte zu erreichen. Der Klausurteil besteht aus vier Seiten.**

**Es sind keine Hilfsmittel zugelassen.**

**Die Klausur besteht aus 4 Seiten.**

Bitte beantworten Sie alle Fragen direkt auf das Angabenblatt.

Nutzen Sie ggf. die Rückseite und kennzeichnen Sie dies entsprechend.

Eigene Schmierblätter sind nicht erlaubt.

Bei mehreren oder mehrdeutigen Lösungen wird die schlechtere Lösung gewertet. Streichen Sie daher ungültige Lösungen eindeutig durch.

**Viel Erfolg!**

**Aufgabe 1:**

(3 Punkte)

Name 3 of the approaches to Information Science mentioned in class.

*Nennen Sie 3 informationswissenschaftliche Herangehensweise, die in Kurs besprochen worden sind.*

**Aufgabe 2:**

(3 Punkte)

In the western world the initial publishing industry can be traced to around the 12th century and was based around: *In der westlichen Welt, die Druckindustrie stammt aus dem 12. Jahrhundert. Von wem wurde sie getrieben?*

- Universities (*Universitäten*)
- Religious institutions (*Religiöse Einrichtungen*)
- Royal or noble families (*Adelsfamilien*)
- Private individuals (*Einzelpersonen*)
- Other \_\_\_\_\_

**Aufgabe 3:**

(5 Punkte)

Calculate the mode, mean, median, IQR and Variance for the following set of numbers. Show your working. *Berechnen Sie den Modus, den arithmetischen Mittelwert, den Median, die IQR, und Varianz für die unten gelisteten Zahlen.*

12	8	5	7	9	7	8	8	5	8
----	---	---	---	---	---	---	---	---	---

Mode	
Mean	
Median	
IQR	
Variance	

**Aufgabe 4:**

(3 Punkte)

When designing experiments one must establish a trade-off between control and [...]

*Wenn man Experimente entwürft, muss man eine Abwägung machen zwischen Kontrolle und [...]*

- Simulation
- Number of participants (*Anzahl der Teilnehmer*)
- Ecological validity (*ökologische Validität*)
- Time to prepare experiments (*Vorbereitungszeit für Experimente*)
- Participant effort (*Mühe der Teilnehmer*)

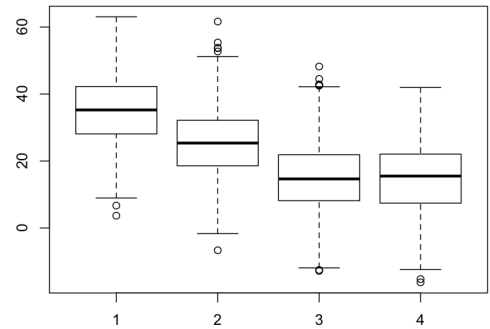
**Aufgabe 5:**

Which of the following of the following distributions has an IQR of 20?

Welche der Verteilungen hat eine IQR von 20?

- 1
- 2
- 3
- 4
- None (keine)

(3 Punkte)



**Aufgabe 6:**

Which of the following are components of systems-based IR experiments (tick all that apply)?

Welcher der folgenden Auswahlmöglichkeiten müssen Teile eines Systems-basierten IR-Experiments sein?

(6 Punkte)

Trifft nicht zu

- Test participants - Probanden
- Retrieval algorithms Retrieval - Algorithmen
- Test collections - Testsammlungen
- Work task scenarios - Arbeitsaufgabenszenarien
- Vector Space Model - das Vektorraummodell
- Topics (queries) - Anfragen

Trifft zu

- 
- 
- 
- 
- 
- 

**Aufgabe 7:**

Mark all of the following which are typical of boolean retrieval.

Welche der folgenden Auswahlmöglichkeiten sind mit boolischem Retrieval assoziiert?

(5 Punkte)

Trifft nicht zu

- Freetext queries. Anfragen in natürlicher Sprache
- Set-based evaluation metrics. Mengen-basierte Evaluationsmetriken
- Boolean logic queries. Anfragen in Form boolischer Logik
- Ranked results list. Sortierte Ergebnislisten
- Binary relevance judgements. Binäre Relevanzbeurteilungen

Trifft zu

- 
- 
- 
- 
- 

**Aufgabe 8:**

A retrieval system returns 10 documents for a query it estimates to be relevant out of a collection of 10,000. 3 of these are actually relevant. 50 relevant documents exist. A user reads half of the returned documents. Estimate Precision, Recall, and F1-Score for this example.

Ein IR-System schätzt von einer Sammlung mit 10,000 Dokumente 10 als relevant ein. 3 der zurückgelieferten Dokumente sind tatsächlich relevant. Es gibt 50 relevante Dokumente in der Sammlung. Der Nutzer liest nur die Hälfte der als relevant geschätzten Dokumente. Berechnen Sie Precision, Recall und F1.

(6 Punkte)

**Aufgabe 9:****(6 Punkte)**

For each of the following levels provide an example of a context factor that would influence how an individual re-finds information.

*Nennen Sie für jede der folgenden Stufen eine Context-Variable, die beeinflussen könnte, wie ein Nutzer Informationen wiederfindet.*

*personal:*

*situational:*

*Socio-economic*

**Aufgabe 10:****(3 Punkte)**

We wish to train a classifier which will predict whether a user will like a given recipe suggestion. Name one approach we looked at in class which would be suitable for this purpose and explain how it would be useful.

*Nennen Sie eine im Kurs besprochene Technik, die benutzt werden könnte, um Rezepte als geeignet oder ungeeignet für einen Nutzer zu klassifizieren. Erklären Sie wie sie nützlich sein k*