



## Information Retrieval Systeme Wintersemester 2014/15 - Teilleistung 1 -

- Beachten Sie zu den Teilleistungen stets die Hinweise im Onlinekurs. Verfolgen Sie hierzu auch die Diskussionen in den Foren des Kurses.
- Geben Sie Ihre Lösung als **ein ZIP-Archiv** ab.
- Die Abgabe erfolgt durch den Upload des ZIP-Archivs über das entsprechende Formular zu dieser Teilleistung im Onlinekurs.
- Benennen Sie das Dokument mit den Lösungen nach dem Schema:

**Nachname-Vorname-TL1.zip**

- Alle Aufgaben ohne Implementierungsaspekte bzw. praktische Anwendung lösen Sie bitte in **EINER zusammenhängenden PDF-Datei**, die Sie mit ins ZIP-Archiv packen.
- Verwenden Sie bei der Programmierung keine weiteren als die angegebenen Bibliotheken. Falls keine Bibliotheken angegeben sind, verwenden wir lediglich die Standard Java-Bibliothek von Oracle.
- Bei der Programmierung bietet es sich an diversen Stellen an mit Exceptions zu arbeiten. „Fehlerfälle“ können Sie sehr einfach behandeln, indem Sie stets eine *RuntimeException* werfen.
- Abgabetermin ist der

**Freitag – 21.11.2014, 23:55 Uhr** (Achtung: es zählt die Serverzeit).

Wir wünschen Ihnen viel Spaß und Erfolg bei der Bearbeitung der Aufgaben!

**Hinweis 1:** Da der Onlinekurs mit halben Punkten nicht umgehen kann, wird es pro Teilleistung 40 Moodle-Punkte geben, die den eigentlichen 4 Punkten pro Teilleistung entsprechen. Am Ende des Kurses werden die Moodle-Punkte Ihrer Teilleistungen addiert, durch 10 dividiert und anschließend auf halbe Punkte gerundet. Sollten Sie z.B. 117 Moodle-Punkte erreicht haben, entspricht dies 11,5 Bonuspunkten. 108 Moodle-Punkte würden bspw. 11 Bonuspunkten entsprechen. Weitere Informationen finden Sie hier: <http://www.uni-bamberg.de/index.php?id=10596>

**Hinweis 2:** Sofern Sie bei der Bearbeitung der Aufgaben auf Schwierigkeiten stoßen bzw. Aufgabenstellungen aus Ihrer Sicht unklar formuliert sind, möchten wir Sie bitten, intensiv und umgehend die Kommunikationsmöglichkeiten über das Forum oder per E-Mail zu nutzen und ggf. nachzufragen. Ein direktes Feedback soll Ihnen und auch Ihren Kommilitoninnen und Kommilitonen den Zugang zu den Aufgaben erleichtern, der aufgrund unterschiedlicher Vorkenntnisse ggf. ausschließlich auf Basis der Aufgabenstellung nicht immer für Alle mit dem gleichen Aufwand möglich sein kann.

### Aufgabe 1 (Grundlegendes, MAP und GMAP, 3 P.)

- Sie wollen einen Evaluations-Korpus für Suchmaschinen erstellen, die auf Beipackzetteln für Medikamente arbeiten. Aus welchen drei Grundbestandteilen muss der Evaluations-Korpus bestehen und wozu könnten Sie in diesem Zusammenhang „Pooling“ einsetzen?
- Im Hinblick auf verschiedene Nutzungsszenarien für IR-Systeme sind Recall und Precision nicht immer gleich wichtig. Hierzu zwei Unterfragen, wobei Sie bitte die Antworten jeweils begründen:
  - Nennen Sie ein Beispiel für ein Anwendungsszenario, in dem Recall besonders wichtig ist.
  - Nennen Sie ein weiteres Beispiel, in dem Precision besonders wichtig ist.
 Überlegen Sie sich zwei Beispiele, die in den Vorlesungsunterlagen noch nicht genannt sind.
- Gegeben seien zwei Systeme A und B, die auf Basis von vier Anfragen miteinander verglichen werden sollen. Die Werte für die durchschnittliche Precision (AP), die beide Systeme dabei erzielt haben, sind in der folgenden Tabelle angegeben:

	System A	System B
Anfrage 1	0,2	0,3
Anfrage 2	0,2	0,25
Anfrage 3	0,1	0,3
Anfrage 4	0,9	0,25

Bestimmen Sie für beide Systeme MAP und GMAP. Runden Sie Ergebnisse auf eine Nachkommastelle.

### Aufgabe 2 (Recall-Precision-Berechnung und Visualisierung, 4 P.)

- Gegeben sei ein IR-System mit dem Namen „IRS-3000“. Sie testen IRS-3000 an zwei Beispielanfragen. Zur ersten dieser Anfragen gibt es 5 relevante Dokumente in der Kollektion und zur zweiten 4.

IRS-3000 liefert für die beiden Anfragen nun entsprechende Rankings. Die folgende Tabelle zeigt, welche der Dokumente im Ergebnis jeweils relevant sind:

Ergebnisse für das System IRS-3000:

Rang	1	2	3	4	5	6	7	8	9	10
Anfrage 1	rel	rel	---	rel	---	rel	---	---	---	rel
Anfrage 2	---	rel	---	rel	rel	---	rel	---	---	---

Berechnen Sie für das System die durchschnittliche Precision an den Standard-Recall-Punkten 0,0; 0,1; 0,2; 0,3 ... bis 1,0 und geben Sie neben dem Ergebnis auch den Rechenweg hierzu an.

- b) Angenommen ein anderes System IRS-2000 hätte folgende AP-Werte an den o.g. Recall-Punkten:

	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
AP	0,75	0,75	0,75	0,7	0,7	0,7	0,7	0,65	0,65	0,65	0,65

Zeichnen Sie beide Systeme IR-2000 und IR-3000 in einen Recall-Precision-Graph. Welches System ist das bessere? Begründen Sie Ihre Antwort.

- c) Zusätzlich soll nun für die Analyse von IRS-3000 eine dritte Anfrage bei der Ermittlung der durchschnittlichen Precision an den Standard-Recall-Punkten mit berücksichtigt werden. Auch hier liefert IRS-3000 ein Ranking von 10 Dokumenten.

Ergebnisse für das System IRS-3000:

Rang	1	2	3	4	5	6	7	8	9	10
Anfrage 1	wie oben									
Anfrage 2	wie oben									
Anfrage 3	---	rel	rel	---	rel	---	rel	rel	---	---

Wir wissen aber, dass es insgesamt 9 relevante Dokumente für Anfrage 3 gibt, von denen IRS-3000 allerdings nur 5 findet. Berechnen Sie nun auf Basis der drei Anfragen die durchschnittliche Precision an den Standard-Recall-Punkten 0,0; 0,1; 0,2; 0,3 ... bis 1,0 für IRS-3000. Dabei können Sie annehmen, dass die Anzahl der relevanten Dokumente im Verhältnis zur Kollektionsgröße verschwindend klein ist.

### Aufgabe 3 (NDCG, 2 P.)

2011

Gegeben seien drei Systeme, mit denen eine bestimmte Anfrage Q bearbeitet wurde. Zu der Anfrage Q sollen in der Dokumentenkollektion insgesamt 7 relevante Dokumente existieren. Die folgende Tabelle zeigt, an welchen Rängen die einzelnen Systeme relevante (rel) und nicht relevante (---) Dokumente im Ergebnis liefern:

Rang	1	2	3	4	5	6	7	8	9	10
System 1	rel	---	rel	---	rel	---	rel	---	---	---
System 2	rel	---	---	rel	---	rel	---	rel	---	rel
System 3	rel	rel	rel	---	---	---	---	---	---	---

- a) Berechnen Sie für die Systeme zur gegebenen Anfrage  $P@3$  und  $P@6$ .  
b) Berechnen Sie für die Systeme die NDCG-Werte bei einer Betrachtung bis Rang 10.

#### Aufgabe 4 (Teststatistiken, 3 P.)

Im Kurs haben Sie drei konkrete Testverfahren kennengelernt. Nennen Sie diese! Berechnen Sie nun für jedes der drei Testverfahren den Wert der Teststatistik. Die folgende Tabelle zeigt dazu die Effektivitätswerte zweier IR-Systeme A und B bei 8 Anfragen. Wir vermuten, dass System B besser als System A ist. Machen Sie bei Ihren Ergebnissen deutlich, welcher Ergebniswert zu welcher Teststatistik gehört.

Hinweis: Es reicht hier aus, jeweils den Wert der Teststatistik zu berechnen. Sie müssen keinen vollständigen Signifikanztest durchführen.

Anfrage	1	2	3	4	5	6	7	8
System A	17	2	15	20	22	50	42	10
System B	28	10	15	18	29	55	49	10

#### Aufgabe 5 (Testdurchführung, 2 P.)

[Achtung: Die Aufgabe ist leichter als sie auf den ersten Blick aussieht!]

Zwei Suchmaschinen (System A und System B) sollen auf Basis von 20 Beispielanfragen miteinander verglichen werden. Die folgende Tabelle zeigt, an welchen Rängen die einzelnen Systeme relevante (rel) und nicht relevante (---) Dokumente im Ergebnis liefern (Rg steht für Rang):

Anfrage	System A					System B				
	Rg 1	Rg 2	Rg 3	Rg 4	Rg 5	Rg 1	Rg 2	Rg 3	Rg 4	Rg 5
1	rel	rel	---	---	---	---	rel	---	---	---
2	---	rel	rel	rel	rel	rel	rel	rel	---	rel
3	rel	---	rel	rel	rel	---	---	rel	---	---
4	rel	---	rel	rel	rel	---	---	---	rel	---
5	---	rel	rel	---	rel	rel	---	---	---	rel
6	rel	---	---	rel	rel	---	rel	rel	---	---
7	rel	---	rel	---	---	---	---	---	rel	---
8	---	rel	---	rel	---	---	---	---	---	---
9	rel	rel	rel	---	---	rel	rel	---	---	rel
10	---	rel	---	rel	---	---	---	---	rel	---
11	---	---	rel	rel	rel	---	---	rel	---	---
12	rel	---	---	rel	---	rel	---	---	---	---
13	rel	rel	---	rel	---	---	---	---	---	---
14	rel	---	rel	rel	rel	rel	---	---	rel	---
15	rel	---	---	---	---	---	---	---	rel	rel
16	rel	---	rel	rel	---	---	rel	rel	---	---
17	rel	---	rel	---	rel	---	---	---	---	---
18	---	rel	rel	---	---	---	---	rel	---	---

19		rel	---	---	rel	---		rel	rel	---	---	rel
20		---	rel	rel	---	rel		---	---	rel	---	---

Der Vergleich soll nun anhand des Effektivitätsmaßes  $P@5$  erfolgen. Dabei soll geprüft werden, ob System A besser ist als System B. Hierzu soll der Vorzeichentest angewendet werden. Erläutern Sie Ihr Vorgehen und gehen Sie dabei auch auf die Sicherheit Ihrer Einschätzung ein.

Die folgende Tabelle gibt den Wert  $B$  der Binomialverteilung für 20 Experimente an:

$k$	$B$	$k$	$B$	$k$	$B$	$k$	$B$
0	0,0000010	6	0,0369644	12	0,1201344	18	0,0001812
1	0,0000191	7	0,0739288	13	0,0739288	19	0,0000191
2	0,0001812	8	0,1201344	14	0,0369644	20	0,0000010
3	0,0010872	9	0,1601791	15	0,0147858		
4	0,0046206	10	0,1761971	16	0,0046206		
5	0,0147858	11	0,1601791	17	0,0010872		

Sofern Sie zur Bearbeitung der Aufgabe Annahmen machen müssen, sind diese zu erläutern.

#### Aufgabe 6 (Recall und Precision, 2 Punkte)

Angenommen Sie hätten ein IR-System entwickelt, das Spam-Mails filtern kann. In einer kleinen Studie haben Sie folgende Werte ermittelt:

	tatsächlich Spam	kein Spam
als Spam erkannt	82	24
nicht als Spam erkannt	31	51

Angenommen Sie wollten auch in diesem Zusammenhang die Maße Recall und Precision verwenden, um die Effektivität des Systems festzustellen. Wie würden Sie Recall und Precision berechnen? Begründen Sie Ihre Berechnung.

#### Aufgabe 7 (MRR und BPREF, 2 P.)

Folgende Tabelle zeigt für zwei IR-Systeme S1 und S2 jeweils das Ergebnisranking im Hinblick auf eine bestimmte Anfrage, zu der es 3 relevante Dokumente (rel) gibt:

Ergebnisranking für die Systeme S1 und S2:

Rang	1	2	3	4	5	6
S1	rel	rel	---	---	---	rel
S2	---	rel	---	rel	rel	---

Berechnen Sie für beide Systeme jeweils:

- den reziproken Rang des ersten relevanten Dokuments
- BPREF

### Aufgabe 8 (Vektorraummodell, 2 Punkte)

ivianu

Gegeben sei die Anfrage „Hund Katze Maus“. Für die drei Begriffe gelten folgende Häufigkeiten:

	Vorkommen in D1	Vorkommen in D2	in Dokumenten der Kollektion
Hund	3	0	40
Katze	2	1	20
Maus	0	1	3

Die Kollektion insgesamt umfasst 100 Dokumente. Bestimmen Sie ein Ergebnisranking der Dokumente D1 und D2 im Hinblick auf die Anfrage „Hund Katze Maus“ und rechnen Sie die entsprechenden Scores aus. Wenden Sie zur Berechnung der Termgewichte und zur Bestimmung der Korrelation die Formeln des Vektorraummodells an. Verwenden Sie den Logarithmus zur Basis 2.

### Aufgabe 9 (Reading Exercise zu TREC, 6 P.)

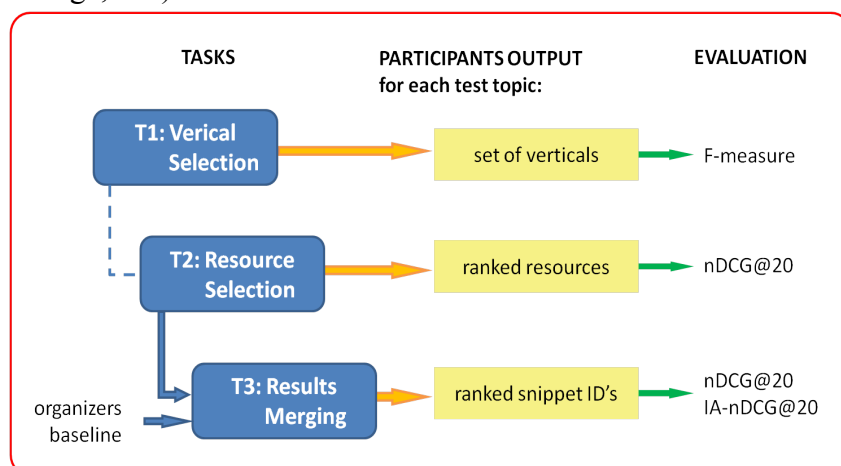
jasch

Ein Bekannter erzählt Ihnen, er habe jüngst ein Poster zum sog. „*TREC Federated Web Search track*“ gesehen. Er hat sich folgende Abbildung abfotografiert, mit der er jedoch nichts weiter anfangen könne. Da Ihr Bekannter die IR Vorlesung besucht hat, kennt er jedoch Konzepte wie *Topics* und *Evaluation* sowie die Evaluationsmaße *F-measure* und *nDCG@20*. Außerdem weiß er, worum es sich grundlegend bei TREC handelt.

Lesen Sie die unten angegebene Quelle und erklären Sie Ihrem Bekannten auf prägnante Weise die Konzepte, die auf der Grafik abgebildet sind und deren Zusammenhänge, sodass der Bekannte nach Lesen Ihres kleinen Artikels alles Wesentliche zum TREC Federated Web Search track verstanden haben sollte. Versuchen Sie dabei präzise zu formulieren.

[Als Längenvorgabe gilt maximal eine Seite A4 in einer 12er Schriftgröße. Bitte nicht falsch verstehen, dies dient in erster Linie dazu, die Aufgabenbearbeitung nicht ausufern zu lassen und Sie vor unnötigem Aufwand zu schützen. Auch kürzere Artikel sind willkommen.]

Im Wesentlichen reicht es bei dieser Aufgabe die genannten Konzepte zu beschreiben (was bedeuten die Bezeichnungen in den Kästen, was bedeuten die Pfeile, was die Wörter am Rande, wie sind die Zusammenhänge, etc.).



Quelle: <https://sites.google.com/site/trecfedweb/>

Anmerkung: Verical Selection sollte eigentlich Vertical Selection lauten.

## Aufgabe 10 (Evaluationsmaße in Java; 14 P.)

ivianus &  
Dima

Im Unterordner *data* des Eclipse-Projekts zu dieser Aufgabe finden Sie drei Dateien (zwei mit der Endung *.trec* und eine mit Endung *.qrel*), die wir im Rahmen einer kleinen Evaluierung verwenden möchten.

Relevanzurteile im IR sind typischerweise in einer QREL-Datei enthalten. Die Informationen darin können zeilenweise knapp beschrieben werden als:

```
<query number> <0> <Document ID> <1 (if is relevant)>
```

Evaluierungsläufe von IR-Systemen werden demgegenüber in sog. TREC-Dateien abgelegt. Eine Zeile darin besteht i.d.R. aus 6 Spalten in der folgenden Formatierung:

```
<query number> <1> <Document ID> <Rank> <Distance from or  
similarity to the query> <Text-Label>
```

In der vorgegebenen Klasse *Evaluation.java* existiert bereits die Methode *List<Set<Integer>> readQueries(File file)* zum Einlesen der QREL-Datei. Alle Dokument IDs zu einer Anfrage werden in einer Menge vom Typ *HashSet<Integer>* gespeichert. Eine Liste erfasst dabei an Position 0, die Menge aller relevanten Dokument IDs zu Anfrage 1; an Position 1, die Menge aller relevanten Dokument IDs zu Anfrage 2, usw.

Implementieren Sie die Methode *evaluate(.)* der Klasse *Evaluation.java* so, dass für eine angegebene TREC-Datei die folgenden Kennzahlen mit Hilfe der Formeln aus den Vorlesungsunterlagen berechnet werden.

- Arithmetisches Mittel der P@10-Werte (*avgP\_at\_10*)
- *MAP*
- *GMAP* (beachten Sie dabei auch die alternativen Berechnungsmöglichkeiten des geometrischen Mittelwerts, z.B. [http://de.wikipedia.org/wiki/Geometrisches\\_Mittel](http://de.wikipedia.org/wiki/Geometrisches_Mittel))
- Arithmetisches Mittel der NDCG-Werte an Rang 10 (*avgNDCG\_at\_10*) (nur unter Berücksichtigung binärer Relevanz)
- Arithmetisches Mittel der BPREF-Werte (*avgBPREF*)

WICHTIG: Dabei soll jede Datei nur einmal gelesen werden. Der Source-Code zum Einlesen, den Sie anpassen können, ist bereits grob vorgegeben.

WICHTIG: Es empfiehlt sich die Implementierung anhand von kleinen selbst erstellten Beispieldateien zu testen!

Ergebniswerte werden stets auf 3 Nachkommastellen gerundet und auf der Konsole ausgegeben. Der Code für die Ausgabe und das Runden ist bereits in den Sourcen enthalten. Daraus ergeben sich auch die Variablen in denen Sie die Kennzahlen erfassen sollen.

Viel Spaß und viel Erfolg beim Bearbeiten der Aufgaben!