

Information Retrieval
Prof. Dr. N. Fuhr

Sascha Kriewel
sascha.kriewel@uni-duisburg.de

Abgabe bis 30. April 2004

Übungsblatt Nr. 1

Aufgabe 1: Mittelwertbildung

In der Vorlesung hast Du gehört, dass es zwei verschiedene Möglichkeiten zur Mittelwertbildung gibt:

Makrobewertung

$$p_M = \frac{1}{N} \sum_{i=1}^N \frac{|REL_i \cap GEF_i|}{|GEF_i|}$$

und Mikrobewertung.

$$p_\mu = \frac{\sum_{i=1}^N |REL_i \cap GEF_i|}{\sum_{i=1}^N |GEF_i|}$$

Stelle entsprechende Formeln für den Recall auf und berechne mit den beiden Ansätzen der Mittelwertbildung über mehrere Anfragen jeweils Recall und Precision für die im Folgenden aufgeführten Anfragen:

Anfrage	$ GEF $	$ REL $	$ REL \cap GEF $
q_1	12	5	3
q_2	120	20	13
q_3	33	50	14
q_4	50	17	8

10 Punkte

Aufgabe 2: Rangordnungen

Gegeben sind zwei IR Systeme 1 und 2. Mit diesen beiden Systemen werden jeweils Anfragen gestartet auf einer Dokumentenkollektion, die bezüglich der Anfrage genau 10 relevante Dokumente enthält.

Die Ergebnisse der beiden Systeme seien wie folgt angegeben:

$$\Delta_1 = (+|-|+|-|-|-|+|+|+|-|-|-|-|+|+)$$
$$\Delta_2 = (-|-|-|+|-|-|-|-|+|-|+|-|+|+|+|-)$$

Bestimme für diese zwei Systeme die Recall- und Precision-Werte nach den einzelnen Dokumenten im Ergebnisranking.

Zeichne dazu ein Recall-Precision-Diagramm mit linearer Interpolation. Kannst Du anhand dieses Graphen eine Aussage über die vergleichsweise Güte der beiden Systeme treffen? Wie würdest Du allgemein das „bessere“ von zwei Retrievalsystemen auswählen?

10 Punkte

Allgemeine Bemerkungen zum Übungsbetrieb

Zur praktischen Übung und Erprobung der Vorlesungsinhalte werden für die Veranstaltung **Information Retrieval** wöchentliche, zweistündige Übungen angeboten. Hierzu werden jeweils in der Vorlesung am Mittwoch Übungsblätter ausgegeben. Zusätzlich stehen die Übungsblätter nach der Vorlesung im WWW zum Download als PDF zur Verfügung:

<http://www.is.informatik.uni-duisburg.de/teaching>

Die bearbeiteten Übungsblätter werden korrigiert und bewertet in der folgenden Woche zurückgegeben. Dazu müssen sie zum jeweils angegebenen Abgabetermin spätestens **bis 13:30** in den entsprechend beschrifteten Briefkasten vor den Räumen LF 135-141 eingeworfen sein. Eine Abgabe der Lösungen per Email ist nicht möglich.

Schreibt bitte auf jedes Blatt Eurer Abgabe Euren Namen, Eure Matrikelnummer und die Aufgabennummer. Ihr könnt die Aufgaben auch in Gruppen zu maximal 3 Studierenden bearbeiten. Dann gehören Namen und Matrikelnummern *aller drei Übungsteilnehmer* mit auf die Abgabe.

Für jede Aufgabe wird die maximal erreichbare Punktzahl auf dem Übungsblatt angegeben. Bei unvollständigen, oder nur teilweise richtigen Lösungen werden entsprechend weniger Punkte vergeben. Im Anschluß an die Vorlesung kann ein Leistungsnachweis in Form einer mündlichen Prüfung erworben werden. Hierzu ist das Erreichen von mindestens 40% der möglichen Übungspunkte, sowie die aktive Teilnahme an den Übungen Voraussetzung. Zur aktiven Teilnahme gehört neben dem regelmäßigen Besuch der Übung auch das Vorführen mindestens einer Aufgabe im Verlauf des Semesters.

Die Übung findet jeweils am **Montag von 8:30-10 Uhr** in Raum **LF 052** statt, erstmals am Montag den **3. 5. 2004**. Für Fragen zur Vorlesung, zu den Übungsaufgaben und zur Bewertung der Abgaben stehe ich im Anschluß an die Übungen, sowie zu meiner Sprechstunde montags von 16-17 Uhr zur Verfügung:

Sascha Kriewel (sascha.kriewel@uni-duisburg.de), Raum LF 137.