

## Übungsblatt 7

17. Dezember 2008

**Hinweis:** Soweit nicht anders angegeben, gibt es für jede korrekt bearbeitete Teilaufgabe einen Punkt. Die Abgabe der Hausübungen ist bis spätestens zum Beginn der nächsten Vorlesung möglich – entweder persönlich direkt vor der Vorlesung oder per Einwurf in den Briefkasten des Instituts (Informatikzentrum, zweiter Stock, vor Raum 238).

### Aufgabe 14 (Maße zur Retrievalbewertung)

- a) Ein IR-System liefert für eine bestimmte Anfrage acht relevante und zehn nicht-relevante Dokumente zurück. Insgesamt gibt es für diese Anfrage zwanzig relevante Dokumente in der Kollektion. In der gesamten Kollektion befinden sich 200 Dokumente. Bestimmen Sie Precision, Recall und Fallout des IR-Systems in diesem Fall.
- b) Eine alternative Definition des F-Measures ist

$$F = \frac{(\beta^2 + 1) \cdot \text{precision} \cdot \text{recall}}{\beta^2 \text{precision} + \text{recall}}.$$

Zeigen Sie, daß die obige Definition für  $\alpha = \frac{1}{\beta^2 + 1}$  äquivalent zur Definition aus der Vorlesung ist.

- c) Für ein bestimmtes Informationsbedürfnis befinden insgesamt vier relevante Dokumente in der gesamten Kollektion. Zwei verschiedene IR-Systeme liefern wie folgt jeweils eine geordnete Liste von zehn Dokumenten als Ergebnis auf eine entsprechende Suchanfrage (der erste Listenplatz steht dabei ganz links; „R“ ist ein relevantes Dokument, „N“ ein nicht-relevantes).

System 1: R N R N N N N N R R  
System 2: N R N N R R R N N N

Bestimmen Sie die Mean Average Precision für jedes der beiden Systeme. Welche Eigenschaften sind für eine Ergebnisliste demnach wichtig, um einen hohen MAP-Wert zu erhalten? (2 Punkte)

- d) Eine Dokumentenkollektion enthält 10 000 Dokumente. Bezüglich einer bestimmten Anfrage sind davon acht relevant. Ein IR-System liefert auf diese Anfrage folgende geordnete Ergebnisliste:

R R N N N N N N R N R N N N R N N N N R

Wie groß sind hier die Precision at 20 und das F-Measure für  $\alpha = \frac{1}{2}$ ? Zeichnen Sie die Precision-Recall-Kurve für diese Ergebnisliste (sowohl interpoliert als auch nicht-interpoliert). (2 Punkte)

### **Aufgabe 15 (Rocchio-Klassifikation)**

Sei  $(d, c)$  ein Dokument/Klasse-Paar aus der Trainingsmenge. Zudem sei  $d'$  ein neues Dokument, das dieselbe Vektorrepräsentation wie  $d$  besitzt. Bei der Rocchio-Klassifikation kann es passieren, daß  $d'$  einer anderen Klasse als  $c$  zugewiesen wird. Geben Sie hierfür ein Beispiel an.  
(2 Punkte)