

Einführung in IR - Winter 2006/07

Dipl.-Inform. Sascha Kriewel, LF 137

sascha.kriewel@uni-due.de

Übungsblatt 4

**Evaluierung
keine Abgabe**

Aufgabe 9: Wiederholung: Recall-Precision-Graphen

(a) Gegeben seien die beiden folgenden linearen Rangordnungen

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

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

(die zugrundeliegende Dokumentenkollektion D_1 enthalte 14 Dokumente und die beiden schwachen Rangordnungen

$$\Delta_3 = (+|+|+|-|-|-|-|+|+|+|-|+|-|-|-|-|-|+|+|- (80))$$

$$\Delta_4 = (+|+|-|+|+|-|-|-|-|-|-|-|-|-|-|+|+|-|-|-|+|- (80))$$

(die zugrundeliegende Dokumentenkollektion D_2 enthalte 100 Dokumente).

Bestimme zunächst auf die übliche Weise *Precision* und *Recall*, und zeichne dazu einen Graphen. Ein Benutzer möchte nun wissen, welche Precision er bei einem Recall von 0,5 erwarten kann. Interpretiere die RP-Graphen entsprechend (auch im Vergleich der Distributionen Δ_1 und Δ_2 bzw. Δ_3 und Δ_4).

(b) Benutze

(i) Probability of Relevance (NR),

(ii) Erwartete Precision (NR), und

(iii) Erwarteter Recall (ND)

um Recall-Precision-Kurven für die obigen Distributionen zu bestimmen. Kann man diese Definitionen auch auf die linearen Rangordnungen anwenden?

Lässt sich die Benutzerfrage nun besser beantworten? Wie viele relevante Dokumente muss der Benutzer im Fall von D_2 finden, um einen Recall von 0,5 zu erhalten? Beeinflusst das die Antwort?