

Übungen zu Internet-Suchmaschinen, Sommersemester 2015

Vu Tran (LF 139)

Sprechstunde nach Vereinbarung

vtran@is.inf.uni-due.de

Übungsblatt 6

keine Abgabe

Aufgabe 12: Retrievalmaße

Gegeben sei eine Kollektion mit 1000 Dokumenten und die folgenden Distributionen:

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

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

Nimm an, dass

- für jede Distribution bzw. dazugehörige Anfrage 50 relevante Dokumente in der Kollektion vorhanden sind
- das System nur Dokumente mit Retrievalwerten¹ $r > 0$ zurückliefert

Berechne für jede dieser Distributionen

- Prec@5
- Recall
- Fallout

Aufgabe 13: Recall und Precision bei schwacher Ordnung

Gegeben ist folgende Distribution. Nimm an, dass hier die gesamte Kollektion beschrieben ist (die also 15 Dokumente enthält):

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

Berechne Recall, Precision und Fallout nach jedem Rang.

Aufgabe 14: (Discounted) Cumulated Gain

Gegeben sei eine Distribution $\Delta_1 = (3|0|1|1|3|2|0)$.

(a) Berechne den Cumulated-Gain-Vektor für Δ_1

¹„Retrieval status value“ bzw. Systemrelevanz

- (b) Berechne den Discounted Cumulated-Gain-Vektor für Δ_1 mit der Abnahmefunktion $d(k) = 1/\log_3 k$ (also $b = 3$).
- (c) Berechne den normalized Cumulated-Gain-Vektor für Δ_1
- (d) Wie kann man (Discounted) Cumulated Gain verwenden, um zwei Systeme zu vergleichen?