Übung zur Vorlesung

Webbasierte Informationssysteme

Heinz Schweppe, Annika Hinze, WS2000/2001

- Übung 2 -

1. Aufgabe

   (a) Angenommen es gibt 10 relevante Dokumente bezüglich der Anfrage q. Berechnen Sie die Recall und Precision - Werte sowie die interpolierten Precision-Werte für folgende Situation: Die Dokumente d5, d9, d1, d7 und d13 werden auf die Anfrage q zurueckgeliefert , und zwar in dieser Reihenfolge. Alle Werte sind verschieden. Relevante Dokumente bezueglich q sind d9 und d7.

   (b) Schlagen Sie Bewertungsmaße für Boolesches Retrieval vor.

2. Aufgabe
   Gesucht sind verschiedene Klassen von Anfragen (mindestens drei). Versuchen Sie die zu definieren und geben Sie für jede Klasse mindestens ein charakteristisches Beispiel an. Welche Bewertungsmaße eignen sich jeweils?
   Hinweis: In der VL wurde kurz von ”Punktanfragen” gesprochen.

3. Aufgabe

   (a) Geben Sie den Gamma- und den Deltacode für folgende Zahlen an: 12, 53, 64

   (b) Dekodieren Sie den folgenden Bitstring, der eine Folge von Gamma-kodierte Zahlen enthält: ”IIIIOOOOIIIIIIOOOOOIIIIIIIOIOIIII”

   (c) Der erste Teil einer Gamma-kodierten Zahl x ist 1 + floor (ld x) in unärer Kodierung. (floor(v) ist ganzzahliger Anteil von v, ld der duale Logarithmus) Wäre es möglich, stattdessen die Kodierung von floor (ld x) zu nehmen? Wenn ja, diskutieren Sie Vor- und Nachteile (Wieviel spart man unter realistischen Annahmen?)

   (d) Schätzen Sie grob die Größe der Lücken in den Dokumentnummern der Postinglisten ab, ab der es sich lohnen würde (Speicherplatz), herkömmliche 32-Bit Integer-Zahlen statt Gamma-kodierter Zahlen zu nehmen.