Übung zur Vorlesung

Webbasierte Informationssysteme

Heinz Schweppe, Annika Hinze, WS2000/2001

- Übung 7 -

1. Aufgabe

   (a) Zur Präfixsuche in einem Suffixfeld benötigt man einen linksseitigen und einen rechtsseitigen Index, der auf die lexikographisch kleinsten bzw. größten Zeichenketten mit Präfix p verweist. Schreiben Sie einen Pseudocodealgorithmus (möglichst funktional) zur Bestimmung des rechtsseitgen Index.

   (b) Kann man (und wenn ja, wie ?) effizient (i) die Anzahl aller Teilstrings p in einem String S, (ii) den längsten sich wiederholenden Teilstring (ii) den längsten sich wiederholenden mit Präfix p, (iii) den am häufigsten auftretenden der Länge k bestimmen, wenn als Datenstruktur verwendet Suffixfelder werden.

   (c) Bauen Sie ein Suffixfeld für die Zeichenkette caacgtcatggtca# nach dem in der VL angegebenen Verfahren auf (siehe ggf. Originalliteratur) und prüfen Sie Ihr Ergebnis durch Konstruktion des entsprechenden Suffixbaums.

   (d) Betrachtet wird das p-Norm Retrieval mit p = 1, p= 2 und p = unendlich (fuzzy) .

      i. Sind die Operationen AND, OR kommutativ? assoziativ?

      ii. Gilt das Distributivgesetz (t1 AND t2) OR t3 = (t1 OR t3) AND (t1 AND t3) ?

2. Projekt-Aufgabe

   (a) Erstellen Sie einen hinreichend grossen Test-Index (bzgl. Anzahl der Mediendateien, Testdateien bzw. Termen, siehe Uebung).

   (b) Implementieren Sie die Anfrageabarbeitung, Suche und Ranking der Antwortmenge. Testen Sie Ihre Implementierung hinsichtlich der Antwortzeit bei Zugriff durch mehrere Benutzer. Schreiben Sie im Bedarfsfall ein Testprogramm zur Simulation von Nutzereingaben. Beruecksichtigen Sie bei Ihrer Auswertung die Indexgroesse.

   (c) Lassen sich bei den Tests in a) Unterschiede zwischen Themen und Medienanfragen feststellen? Begruenden Sie.