Übung zur Vorlesung

Webbasierte Informationssysteme

Heinz Schweppe, Annika Hinze, WS2000/2001

- Übung 3 -

1. Aufgabe

   (a) Kodieren Sie mit dem Golomb-Code (b = 7) die Zahlen 7, 8 und 13

   (b) Dekodieren Sie die Bitfolge: "IIIOOIIIOIOIOO1" (b = 6)

2. Aufgabe

   (a) Geben sie die Codewortlänge für die natürliche Zahl x im Delta-Code an

   (b) Leiten Sie daraus die implizit angenommene Wahrscheinlichkeit für die Lückengröße x ab. Hinweis: verwenden Sie l(x) = - ld p[x] , also die Aussage, dass die optimale Codelänge der negative Logarithmus der Wahrscheinlichkeit des Vorkommens von x ist.

3. Aufgabe (insgesamt 3 Wochen Bearbeitungszeit)
   Entwerfen, Implementieren und Testen Sie einen Index, der innerhalb der im Projekt entwickelten Suchmaschine eingesetzt werden koennte. Ziel der Implementierung ist eine schnelle Abarbeitung von Anfragen. Die Aufgabe wird ueber einen Zeitraum von drei Wochen bearbeitet, die einzelnen Zwischenschritte in den Uebungen besprochen:

   (a) Entwickeln Sie eine geeignete Datenstruktur zum Speichern aller fuer das Projekt benoetigten Daten; diese werden in der Uebung am 8.11. besprochen. Legen Sie den geplanten Ablauf von Indexaufbau und Abfrageabarbeitung auf Basis der von ihnen gewaehlten Datenstruktur dar.
   Abgabe: 15.11.2000

   (b) Implementieren Sie die Indexstruktur und die Indexierung. Nutzen Sie dazu als Datengrundlage die Sammlung von Datenfiles, die aus den man-pages erstellt wurde (Link auf der Homepage der VL). Implementieren Sie in Java, ermoeglichen Sie die Einbindung ihres Indexierungsmoduls in einen groesseren Projektkontext durch einen modularen Aufbau der einzelnen Komponenten und eine gute Dokumentation des Quellcodes und der Schnittstellen. Die Inverslisten sollen mit einem in der VL vorgestellten Kompressionscode ihrer Wahl kodiert werden (z.B. Gamma oder Golomb-Code).
   Abgabe: 22/29.11.2000

   (c) Erarbeiten Sie geeignete Kriterien und Test-Strategien zur Evaluation des Index und testen Sie ihre Implementierung.
   Abgabe: 22/29.11.2000