

Algorithmen und Datenstrukturen II

SoSe 2011, **3. Aufgabenblatt**, Abgabe 01.06.2011

Gegeben seien die Zeichenketten $S_1 = \text{EIN_NEGER_MIT_GAZELLE_ZAGT_IM_REGEN_NIE}$
und $S_2 = \text{ALLE_AALE_MALEN}$
für Aufgaben 9 und 10.

Aufgabe 9

14 Punkte

- Zählen Sie separat in S_1 und S_2 die Häufigkeiten der Zeichen und geben Sie diese in je einer Tabelle für S_1 und S_2 an. Versichern Sie Sich, dass Ihre Zählung stimmt, damit Sie die beiden folgenden Teilaufgaben auch richtig lösen können. (2 Punkte)
- Konstruieren Sie anhand der in S_1 und S_2 ermittelten Häufigkeiten die entsprechenden Huffman-Codierungs-Bäume T_1 und T_2 . Gibt es mehr als zwei Bäume mit minimaler Häufigkeit, so werden diejenigen ausgewählt, die die am frühesten im Alphabet erscheinenden Zeichen enthalten. Die Zuordnung als linker und rechter Teilbaum bestimmt sich ebenfalls alphabetisch: der Baum mit dem früher vorkommenden Zeichen wird linker Teilbaum. Das Leerzeichen `_` kommt alphabetisch vor `A`. (6 Punkte)
- Geben Sie die Code-Werte jedes Zeiches für T_1 und T_2 an. Kodieren Sie die Zeichenkette S_1 mit T_1 ; und kodieren Sie die Zeichenkette S_2 sowohl mit T_1 als auch mit T_2 . Was fällt Ihnen auf, wenn Sie beide Kodierungen von S_2 vergleichen? (6 Punkte)

Aufgabe 10

6 Punkte

- (3 Punkte) Betrachten Sie die Zeichenkette S'_2 , die aus S_2 durch Entfernen aller Leerzeichen `'_'` entsteht. Wenden Sie die Burrows-Wheeler-Transformation auf S'_2 an. Alle Zähler beginnen bei 0. Geben Sie die alphabetisch sortierte Matrix der Rotationen von S'_2 und die Ausgabe der Transformation (letzte Spalte und Zeilennummer des Originalblocks) an.
- (3 Punkte) Führen Sie die Rücktransformation durch (vgl. Vorlesung). Hierbei ist es ausreichend, den Gesamtpfad zu zeichnen und Start- und Endpunkte zu markieren.

Aufgabe 11

6 Punkte

Gegeben sei die Zeichenkette $T = \text{BIBIS_STARKBIER_IST_BISTABIL}$

Wenden Sie den Algorithmus LZ77 (Sliding Window Lempel-Ziv ohne Optimierungen, laut Vorlesung) auf T an. Der Buffer fasse 4 Zeichen. Geben Sie für je einen Lauf mit Dictionary-Länge 10 und 20 die Ausgabe des LZ77 an.

Aufgabe 12

6 Punkte

Gegeben seien die Zeichenketten $U = \text{RAUMKLANGTRAUMSAUSE}$
und $Q = \text{AUAUA}$

Wenden Sie den Algorithmus von Knuth-Morris-Pratt an, um in U nach Q zu suchen. Geben Sie dazu zuerst die next-Tabelle für Q an. Notieren Sie dann die Folge der Indexpaare (i, j) , für die $U[i]$ mit $Q[j]$ verglichen wird.