

Algorithmen und Datenstrukturen II

SoSe 2010, **3. Aufgabenblatt**, Abgabe 26.05.2010

Gegeben seien die Zeichenketten $S_1 = \text{HASEN_ABER_HABEN_ALLE_ALLELE}$
und $S_2 = \text{ANANASSAHNE_AN_NASE}$
für Aufgaben 7, 8, und 9.

Aufgabe 7

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 einen Baum mit minimaler Häufigkeit, so wird jeweils derjenige ausgewählt, der das am frühesten im Alphabet erscheinende Zeichen enthält. Das Leerzeichen `_` kommt alphabetisch vor `A`. (6 Punkte)
- 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 . (6 Punkte)

Aufgabe 8

8 Punkte

Wenden Sie den Algorithmus LZ77 (Sliding Window Lempel-Ziv ohne Optimierungen, laut Vorlesung) auf S_1 und auf S_2 an. Der Buffer fasse 4 Zeichen. Geben Sie die erzeugte Ausgabe bei Längen 5 und 15 des Dictionary jeweils für die Zeichenketten S_1 und S_2 an. Sie müssen also insgesamt vier Läufe des LZ77 durchführen. (8 Punkte)

Aufgabe 9

8 Punkte

Wenden Sie die Burrows-Wheeler-Transformation auf S_1 und S_2 an. Geben Sie als Zwischenergebnis die Sortierung der Teilstrings an, entweder in Form der Matrix mit allen rotierten Versionen der Zeichenketten nach alphabetischer Sortierung (vgl. Vorlesung) oder (platzsparender), durch Angabe des Sortierranges bei jedem Zeichen in den Original-Zeichenketten S_1 und S_2 . Geben Sie als Endergebnis die permutierten Zeichenketten und die Zeilennummer des Originalblocks an. Die Zeilennummerierung beginne bei 0. (8 Punkte)