Datenkompression: Übungsblatt 1

Enno Ohlebusch
Timo Beller

Das Übungsblatt wird am 30.04.2013 besprochen.

Aufgabe 1

Betrachten Sie die folgenden Mengen von Codewörtern. Welche der Mengen sind Bildbereich eines eindeutig dekodierbaren Codes und warum? Welche verletzen die McMillan'sche Ungleichung?

1. $\{0, 01, 11, 111\}$
2. $\{1, 00, 010, 011\}$
3. $\{00, 10, 11, 001, 101\}$
4. $\{00, 11, 010, 0100\}$

Aufgabe 2

Sei $M = \{a, b, c, d\}$ und l_1, l_2 die Längenfunktionen mit $l_1(a) = 3, l_1(b) = 3, l_1(c) = 3, l_1(d) = 2$ bzw. $l_2(a) = 5, l_2(b) = 1, l_2(c) = 4, l_2(d) = 3$. Konstruieren Sie Präfixcodes mit den entsprechenden Längen für M.

Aufgabe 3

Zeigen Sie, dass die Partitionierung gemäß dem Shanon-Fano-Algorithmus nicht immer optimal ist. D.h. es kann eine andere Partitionierung geben, bei der die Differenz zwischen den Summen der Häufigkeiten kleiner ist. Liefert eine Partitionierung mit minimalem Unterschied immer einen besseren Code? Welchen Nachteil hätte es, wenn der Shanon-Fano-Algorithmus eine Partitionierung mit minimalem Unterschied fordern würde?

Aufgabe 4

Schreiben Sie ein Programm, dass für einen Eingabestring die mittlere Codewortlänge der Shanon-Codierung, der Shanon-Fano-Codierung und der Huffman-Codierung ausgibt.

Aufgabe 5

Ein Eingabestring mit $\sigma = 256$ verschiedenen Symbolen soll mittels Huffman-Code codiert werden. Ist es möglich, dass der Huffman-Code einem Symbol ein Codewort zuordnet, dass 255 Zeichen lang ist? Wenn dies möglich ist, wie lang muss die zugehörige Eingabe dann mindestens sein?