

Arbeitsblatt 3: Verlustfreie Komprimierung von Texten

Fragestellung

Wie können Texte komprimiert gespeichert werden?

Benötigte Materialien

- Browser oder Kopiervorlagen
- ASCII Tabelle
- Elektronische Beilagen

Beschreibung der Aufgabe

Ein Sprichwort aus Äthiopien soll möglichst kompakt gespeichert werden:

Je kleiner die Eidechse, umso grösser ist ihre Hoffnung, ein Krokodil zu werden.

Mit Hilfe einer ASCII-Tabelle kann jedes Zeichen aus dem Text mit acht Einsen oder Nullen codiert werden. Das Sprichwort bestehend aus 80 Zeichen, kann wie folgt in ASCII codiert werden.

```
01001010 01100101 00100000 01101011 01101100 01100101 01101001 01101110
01100101 01110010 00100000 01100100 01101001 01100101 00100000 01000101
01101001 01100100 01100101 01100011 01101000 01110011 01100101 00101100
00100000 01110101 01101101 01110011 01101111 00100000 01100111 01110010
11110110 01110011 01110011 01100101 01110010 00100000 01101001 01110011
01110100 00100000 01101001 01101000 01110010 01100101 00100000 01001000
01101111 01100110 01100110 01101110 01110101 01101110 01100111 00101100
00100000 01100101 01101001 01101110 00100000 01001011 01110010 01101111
01101011 01101111 01100100 01101001 01101100 00100000 01111010 01110101
00100000 01110111 01100101 01110010 01100100 01100101 01101110 00101110
```

Der Code besteht aus 80 x 8 Einsen und Nullen, also aus 640 Bit. David A. Huffman entwickelte 1952 eine raffinierte Methode, welche Texte viel kompakter speichern kann. Anstelle einer ASCII Tabelle, wird eine spezielle Codierungstabelle für jeden Text erstellt. Die Zeichen welche häufig vorkommen werden mit einem kurzen Code beschrieben.

Die elektronische Beilage kann zu jedem gewählten Text eine Hoffman Tabelle, respektive einen Hoffman Baum erzeugen und Texte mit Hilfe der Hoffman Methode komprimieren.

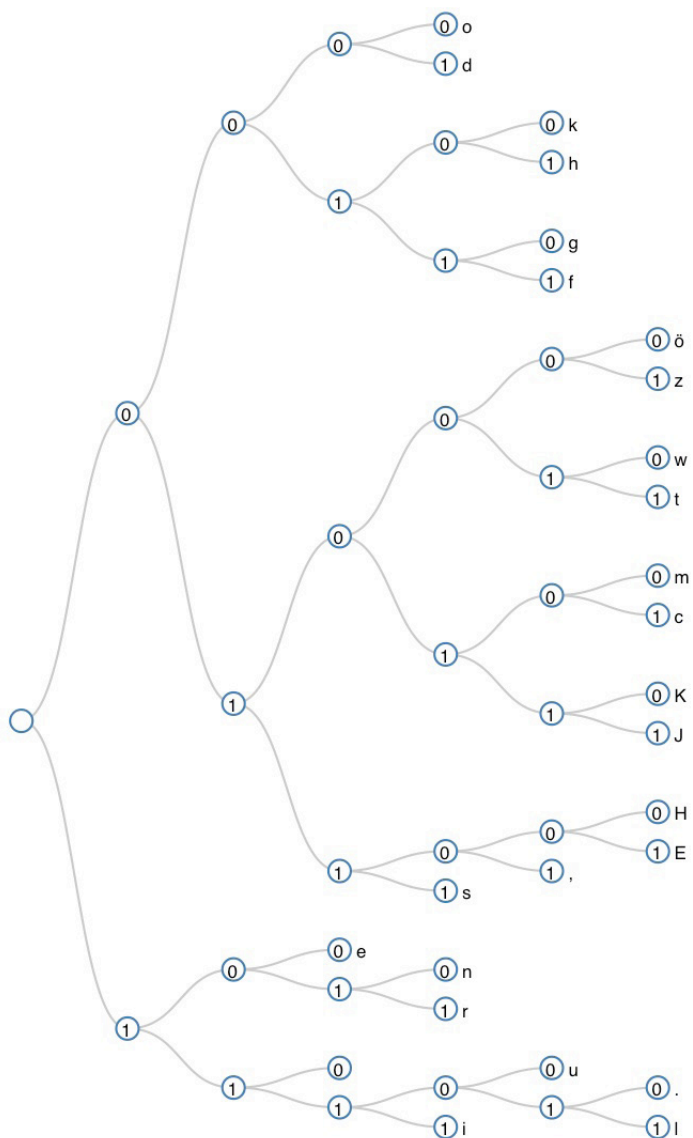
Das äthiopische Sprichwort kann durch folgenden Code kompakter gespeichert werden:

```
010111100110001001110111001111101010010111100001111110011001100111110001100
010101001010111100011011101110001010001110000110001101011010000011101111001-
011110111101110100111101111001011011100110011000000000111001111010111001010
00110011011101001111101011001011010110000001000000000111111101111001000111-
100110010010100101100011001010111010
```

Anstelle der 640 Nullen und Einsen kann das Sprichwort mit 336 Einsen und Nullen gespeichert werden. Damit kann fast die Hälfte an Speicherplatz gespart werden

Decodierung des Sprichworts

Zur Decodierung wird der folgende Huffman-Baum verwendet.



Bei der Decodierung wird bei der Wurzel links begonnen, kommt eine Eins folgt man dem Zweig nach unten, bei einer Null folgt man dem Zweig nach oben, bis am Ende ein Zeichen oder Buchstaben steht.

Der ersten Teil des Codes **0100101001100011101010** wird wie folgt entschlüsselt:

010111 = J

100 = e

110 = (Leerzeichen)

00100 = k

111011 = l

Aufgabe 1: Decodieren mit Hilfe des Huffman Baums

Zu den folgenden Codes sollen die entsprechenden Zeichen gefunden werden. Benutze dazu den Huffman-Baum auf der letzten Seite.

011000

1111

111011

00111

100

Aufgabe 2: Eigenen Text codieren

Nutze die elektronische Beilage um einen eigenen Text zu codieren.

Miss wie viele Bits (Nullen und Einsen) mit der ASCII Codierung und wie viele mit der Huffman Codierung benötigt werden.

Um welchen Faktor kann der Text komprimiert werden?

Die folgende Abbildung zeigt eine Huffman-Codierung mit Hilfe der elektronischen Beilage.

