

Übungen zu Einführung in die Informatik II

Aufgabe 1 **Eigenschaften der Entropie**

- a) Die Entropie ist bekanntlich durch

$$H(s) = - \sum_{i=1}^q p_i \text{ld } p_i$$

gegeben; außerdem ist $0 \leq p_i \leq 1 \Rightarrow \text{ld } p_i \leq 0$. Somit besteht die Summe nicht aus negativen Summanden und kann insgesamt nicht negativ sein.

- b)

$$\begin{aligned} H(S) = 0 &\Leftrightarrow p_i \text{ld } p_i = 0, \forall i \in \{1, \dots, q\} \\ &\Leftrightarrow p_i = 0 \vee (\text{ld } p_i = 0), \forall i \in \{1, \dots, q\} \\ &\Leftrightarrow p_i = 0 \vee p_i = 1, \forall i \in \{1, \dots, q\} \end{aligned}$$

Da aber $\sum_i p_i = 1$ sein muß, muß genau ein $p_i = 1$ sein und das Prädikat $\forall j : j \neq i \Rightarrow p_j = 0$ gelten. Die Entropie erreicht in diesem Fall also ihr Minimum.

Bemerkung: $\lim_{x \rightarrow 0} x \cdot \log x$ konvergiert nach der Regel von l'Hospital gegen 0.

- c) In dem angegebenen Satz wählen wir in dem angegebenen Hilfssatz $x_i = p_i$ und $y_i = 1/q$. Dann folgt:

$$\begin{aligned} H(S) &= \sum_{i=1}^q p_i \text{ld } \frac{1}{p_i} \\ &\leq \sum_{i=1}^q p_i \text{ld } q \\ &= \text{ld } q \sum_{i=1}^q p_i \\ &= \text{ld } q \end{aligned}$$

Bemerkung: Der angegebene Hilfsatz läßt sich mit der bekannten Abschätzung $\ln x \leq x - 1$, für $x > 0$, leicht beweisen.

- d) Hinweis: Die Aufgabenstellung sollte lediglich die Richtung

$$H(S) = \text{ld } q \Leftrightarrow \forall i \in \{1, \dots, q\} : p_i = \frac{1}{q}$$

enthalten.

Einsetzen der Gleichverteilung $p_i = 1/q$ in die Definition der Entropie liefert: $H(S) = \text{ld } q$ als maximale Entropie und damit die Behauptung.

Aufgabe 2 (Lösungsvorschlag)

Die ersten 15 Kontonummern lauten:

- | | | |
|---------|----------|----------|
| 1) 1000 | 5) 0333 | 12) 1113 |
| 2) 0011 | 6) 0444 | 13) 1231 |
| 3) 0102 | 7) ... | 14) 1322 |
| 4) 0220 | 11) 0999 | 15) 1455 |

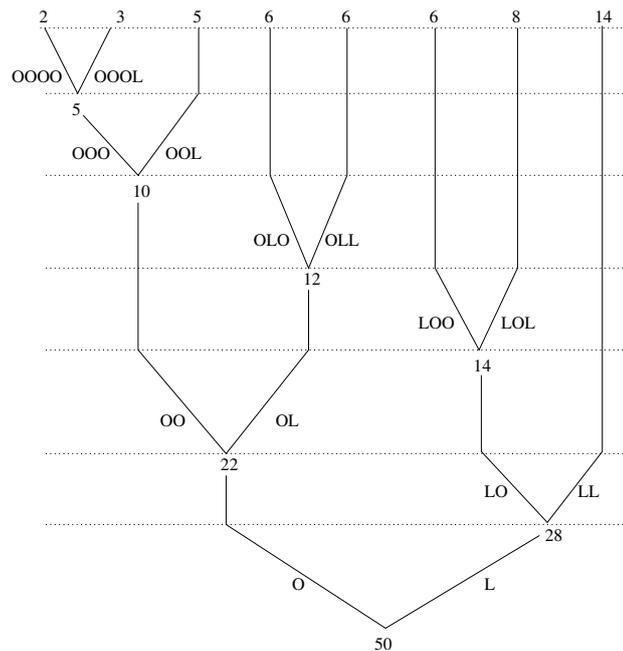
Ein Beispiel für eine falsche Kontonummer: 1327; Korrektur: 1322

Aufgabe 3 Huffman - Algorithmus (Lösungsvorschlag)

a) Die Ableitung der Codierung des Alphabet Z mit der Wahrscheinlichkeitsverteilung

z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8
2	3	5	6	6	6	8	14

veranschaulichen wir mit Hilfe des folgenden Baumdiagramms:



Vorgehensweise:

- Aufbau des Baumdiagramms von oben nach unten, wobei jeweils die Zeichen mit minimalem $p_{z'} + p_{z''}$ zu einem neuen Zeichen z_{neu} mit der Wahrscheinlichkeit p_{neu} zusammengefaßt werden. Die gestrichelten Linien symbolisieren dabei die verschiedenen rekursiven Schritte.
- Ist die Wurzel des Baumes erreicht, so durchläuft man den Baum von unten nach oben.
- Hierbei annotieren wir die Kanten des Graphen mit der Codierung, der im jeweiligen Rekursionsschritt vorhandenen Zeichen;
- Tritt bei Traversieren des Baumes ein Verzweigung auf, so hängen wir im linken Ast an das Ende der bisherigen Codierung ein O (bzw. ein L) und im rechten Ast ein L (bzw. ein O) an.

Somit ergibt sich unter anderem folgende Codierung:

$c(z_1)$	$c(z_2)$	$c(z_3)$	$c(z_4)$	$c(z_5)$	$c(z_6)$	$c(z_7)$	$c(z_8)$
0000	000L	00L	OLO	OLL	LOO	LOL	LL

b) Die Fano-Bedingung besagt bekanntlich, dass keine Codierung eines Zeichens z ein Präfix der Codierung eines Zeichens $z' \neq z$ ist; (Man nennt ein Wort s' Präfix eines Wortes s , wenn ein Wort x existiert, so dass $s = x \circ s'$.) Man erkennt bereits an der Definition des Huffman-Algorithmus, dass die Zeichenmenge in jedem Rekrusionsschritt in drei Teilmengen aufgeteilt wird und diese Teilmengen jeweils von rechts nach links unterschiedlich codiert werden. Somit kann keine Codierung Präfix einer anderen Codierung sein.

c) Die mittlere Länge L der Codierung eines Zeichens ist bekanntlich durch

$$L = \sum_{z \in Z} (p_z \cdot |l(z)|)$$

gegeben, wobei $l(z)$ die Länge der Codierung des Zeichens z bezeichnet. Für obiges Alphabet erhalten wir:

$c(z_1)$	$c(z_2)$	$c(z_3)$	$c(z_4)$	$c(z_5)$	$c(z_6)$	$c(z_7)$	$c(z_8)$
0000	000L	00L	OLO	OLL	LOO	LOL	LL
4	4	3	3	3	3	3	2

Somit ergibt sich $L = ???$

Aufgabe 4 Morse-Code (Lösungsvorschlag)

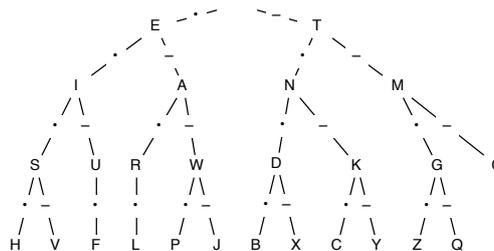
Im Jahr 1884 sendete Samuel Morse eine Nachricht mit einem Fernschreiber von Washington nach Baltimore. Er verwendete dazu einen Code der heute als *Morse-Code* bekannt ist. Dieser Code besteht aus einer Folge von langen und kurzen Tönen.

a) Suchen Sie im Internet nach einer Code-Tabelle für den Morse-Code.

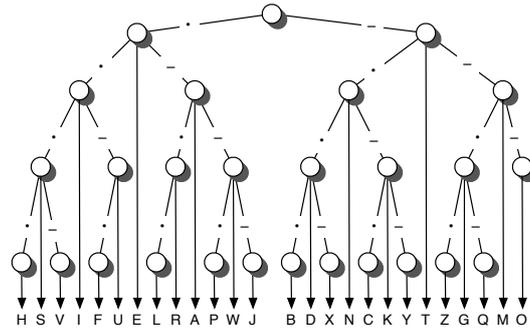
b) Was ist das Alphabet beim Morse-Code?

Das Alphabet ist $\{-, \cdot\}$, also Punkt und Strich; eigentlich kommt aber noch ein drittes Zeihe dazu, das Leerzeichen zwischen zwei Buchstaben

c) Untersuchen Sie den Morse-Code auf Einhaltung der Fano-Bedingung. Erstellen Sie dazu den kompletten Code-Baum für den Morse-Code.



Der Morse-Code erfüllt die Fano-Bedingung nicht. Beispiel: e entspricht \cdot , r entspricht $\cdot - \cdot$. Nimmt man jedoch das Leerzeichen als Alphabetsymbol hinzu, so ergibt sich folgender Code-Baum:



In diesem Code-Baum ist die Fano-Bedingung erfüllt, damit ist der Morse-Code eindeutig decodierbar.

d) Warum sind die Code-Sequenzen unterschiedlich lang?

Kurze Sequenzen entsprechen häufigen Buchstaben, lange Sequenzen entsprechen selteneren Buchstaben, dadurch wird die Länge des codierten Textes kurz. Samuel Morse ging in eine Schriftsetzerei in Philadelphia und zählte dort die Buchstaben in der Schublade eines Schriftsetzers. Er fand dort 12000 mal den Buchstaben e, 9000 mal t, 8000 mal die Buchstaben a, i, n und m. Dies spiegelt sich in der Länge der Sequenzen für die einzelnen Buchstaben wieder.