

MOTIVATION

a) Untersuche den Morsecode auf die Länge der Buchstabencodierung.

a	.	-	b	-	.	.	.	c	-	.	-	.	d	-	.	.
e	.		f	.	.	-	.	g	-	-	.		h	.	.	.
i	.	.	j	.	-	-	-	k	-	.	-		l	.	-	.
m	-	-	n	-	.			o	-	-	-		p	.	-	-
q	-	-	.	-	.			s	.	.	.		t	-		
u	.	.	-	.		-	-	v	.	.	.	-	w	.	-	-
y	-	.	-	-	.	.		z	-	-	.	.	1	.	-	-
3	.	.	.	-	-	.	-	4	.	.	.	-	5	.	.	.
7	-	-		8	-	-	-	.	9	-	-	-
												.	0	-	-	-

Zwischen zwei Zeichen ist immer eine kurze Pause. Wäre eine gemorste Zeichenkette ohne Pausen noch eindeutig? Finde Beispiele.

b) Wird das Zeichen A mit der Bitfolge „10“, das Zeichen B mit der Folge „01“ und C mit „0“ codiert, dann wird die Zeichenkette ABC zu „10010“.

Diese Folge wird aber auch von der Zeichenkette „ACA“ erzeugt. Es ist also nicht mehr eindeutig erkennbar, für welche Zeichenfolge die Bitfolge „10010“ steht.

Finde eine Codierung für die Buchstaben A, B und C, so dass alle Zeichenketten eindeutig kodiert werden können.

c) Eine Textdatei, deren Inhalt die Zeichenkette **alea_lacta_est** ist, soll mit der folgenden Buchstabencodierung codiert und komprimiert werden:

a 00 l 0100 e 011 i 0101 c 100 t 101 s 111 _ 110

Codiere den Satz. Untersuche, welche Kriterien (die selbst gefunden werden müssen) die Buchstabencodierung erfüllt.

Zusammenfassung

Verhindern von Mehrdeutigkeiten

Um Mehrdeutigkeit zu verhindern, müssen die den Zeichen zugewiesenen Bitfolgen **präfixfrei** sein, d. h. keine Bitfolge, die für ein einzelnes Zeichen steht, darf am Anfang eines anderen Zeichens enthalten sein. (Quelle: wikipedia.de)

Entropiecodierung

Häufig auftretende Elemente (z.B. Buchstaben) werden der Häufigkeit nach mit weniger Zeichen versehen als Elemente (Buchstaben), die seltener benutzt werden. Die Entropiecodierung ist verlustfrei.

Als bekanntes Verfahren verwendet das "**zip-Dateiformat**" unter anderem eine Entropiecodierung auf der letzten Codierungsstufe.

Huffman-Codierung (nach David Huffman, 1952)

Beispiel: alea_iacta_est ("Der Würfel ist gefallen")

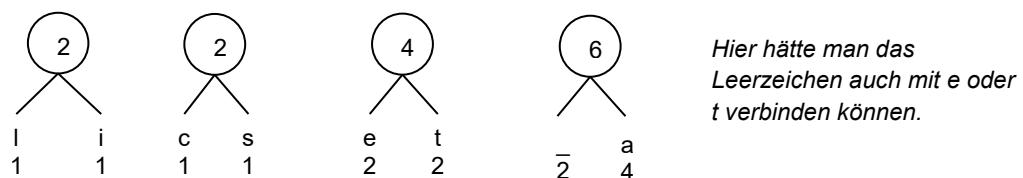
Schritt 1

Zähle die Buchstabenhäufigkeit (auch das Leerzeichen "_") und schreibe sie unter die einzelnen Buchstaben.

a	1	e	i	c	t	s
4	1	2	1	1	2	1
						2

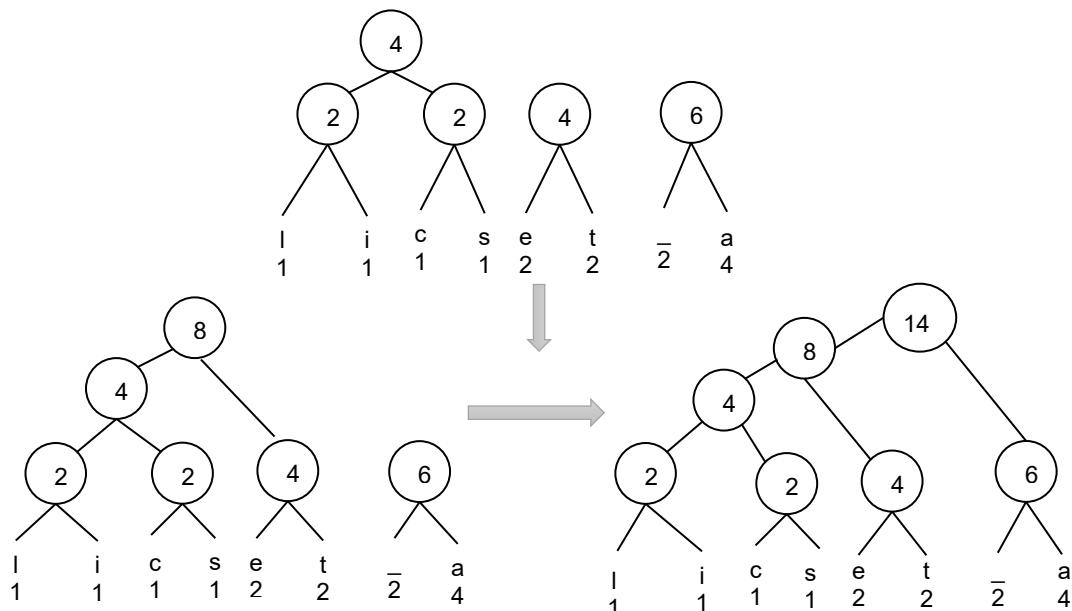
Schritt 2

Ordne die Buchstaben der Häufigkeit nach und verbinde die jeweils beiden kleinsten Buchstabenhäufigkeiten miteinander. In den Knoten kommt die Summe der Buchstabenhäufigkeiten.



Schritt 3

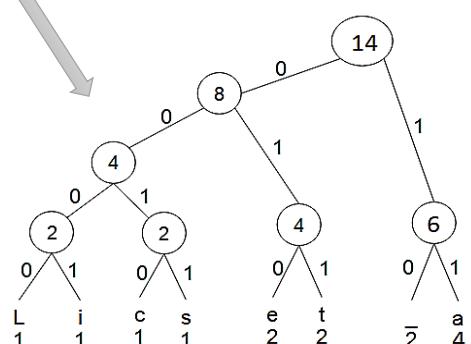
Es werden immer weiter die beiden kleinsten Buchstabenhäufigkeiten entweder unter den Zahlen oder in den Knoten herausgesucht und verbunden, bis ein Baum entsteht.



Schritt 4

Nach der Erstellung des Baumes werden alle Kanten nach links mit 0 und alle Kanten nach rechts mit 1 beschriftet. Der Pfad zu einem Buchstaben ergibt die Codierung.

l 0000	i 0001
c 0010	s 0011
e 010	t 011
_ 10	a 11



Aufgaben

Aufgabe 1

Codiere mit der *Huffman-Codierung* das Wort *a b r a c a d a b r a*.

Schreibe die Bitfolge auf und berechne den Kompressionsfaktor gegenüber einer ASCII-Codierung.

Aufgabe 2

Ein Satz wurde über eine Datenleitung komprimiert übertragen:

0110110111010100011101110001011001001101111011100

Zudem wurde die folgende Buchstabentabelle mit übertragen (oder vorher vereinbart):

00	n	0110	d	1000	r
0111	m	1001	t	101	g
11	e	010	Leerzeichen	/	_

a) Decodiere den Satz.

b) Beweise, dass die Codierung eindeutig ist und es aufgrund der fehlenden Worttrennzeichen keine Codierungsfehler geben kann.

Aufgabe 3

„Ostsee“ wurde mit folgendem Code komprimiert: 0: 0 t: 00 e: 01 s: 1
Begründe, dass der Code nicht eindeutig (präfixfrei) ist.

Aufgabe 4

Wie kann man zeigen, dass der Morsecode nicht präfixfrei ist?