Karlsruher Institut für Technologie

Institute of Information Security and Dependability
Dr. Mattias Ulbrich
https://formal.kastel.kit.edu/

Übungsblatt 6

Grundbegriffe der Informatik — Winter 2023/24

Tutor*in: Tutorium Nr.:

Nach-,Vorname 1: ,

Matr.nr. 1:

Nach-,Vorname 2: ,

Matr.nr. 2:

Ausgabe: 5. Dezember 2023, 14:30 Uhr

Abgabe: 15. Dezember 2023, 12:30 Uhr

Bitte beachten Sie die Hinweise auf der letzten Seite.

---

*Von Tutor*in auszufüllen:*

Blatt 6: / 21

Blätter 1 – 6, Stud. 1: / 122

Blätter 1 – 6, Stud. 2: / 122

Aufgabe 1 - Entschlüsselung (6 Punkte)

Gegeben seien die Alphabete $A = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}\}$ und $B = \{\mathtt{0}, \mathtt{1}\}$ sowie die Funktionen $f\colon A \to B^*$, $g\colon A \to B^*$ und $h\colon A \to B^*$, die gemäß der folgenden Tabelle definiert sind. Den Funktionswert $h(d)$ sollen Sie in Teilaufgabe (e) angeben.

x	a	b	c	d
$f(x)$	00	10	01	11
$g(x)$	00	10	01	011
$h(x)$	00	10	01	? ? ?

Dazu sei $f^{**}\colon A^* \to B^*$ der von f und $g^{**}\colon A^* \to B^*$ der von g induzierte Homomorphismus.

a) Geben Sie $f^{**}(\mathtt{cbd})$, $g^{**}(\mathtt{cba})$ und $g^{**}(\mathtt{da})$ an. (1 Punkt)

b) Sind f^{**} und g^{**} ε-frei? Sind sie präfixfrei? Begründen Sie jeweils kurz Ihre Antwort. (1 Punkt)

c) Geben Sie eine Decodierung $l_f\colon B^* \to (A \cup \{\perp\})^*$ von f^{**} an. (1 Punkt)

d) Geben Sie eine Decodierung $l_g\colon B^* \to (A \cup \{\perp\})^*$ von g^{**} an. (2 Punkte)

e) Ergänzen Sie die Definition von $h\colon A \to B^*$ in der Tabelle oben, so dass der von h induzierte Homomorphismus h^{**} *keine* Decodierung besitzt. Begründen Sie, warum es keine Decodierung gibt. (1 Punkt)

Lösung 1

a) $f^{**}(\mathtt{cbd}) = \mathtt{011011}$ $\quad g^{**}(\mathtt{cba}) = \mathtt{011000}$ $\quad g^{**}(\mathtt{da}) = \mathtt{01100}$

b) Sowohl f^{**} als auch g^{**} sind ε-frei, denn f und g bilden kein Element von A auf ε ab. Damit ist $f^{**}(w) = \varepsilon$ bzw. $g^{**}(w) = \varepsilon$ genau dann, wenn $w = \varepsilon$.

f^{**} ist präfixfrei, denn alle Elemente aus A werden von f auf Wörter der Länge 2 abgebildet. Also gilt für alle $x_1, x_2 \in A$, dass, wenn $f(x_1)$ ein Präfix von $f(x_2)$ ist, $f(x_1) = f(x_2)$ gelten muss, also $x_1 = x_2$.

g^{**} ist nicht präfixfrei, denn $g^{**}(\mathtt{c}) = \mathtt{01}$ ist ein Präfix von $g^{**}(\mathtt{d}) = \mathtt{011}$.

c)

$$l_f\colon B^* \to (A \cup \{\mathtt{\perp}\})^*$$
$$w \mapsto \begin{cases} \mathtt{a} \cdot l_f(w'), & \text{falls } w = \mathtt{00} \cdot w' \\ \mathtt{b} \cdot l_f(w'), & \text{falls } w = \mathtt{10} \cdot w' \\ \mathtt{c} \cdot l_f(w'), & \text{falls } w = \mathtt{01} \cdot w' \\ \mathtt{d} \cdot l_f(w'), & \text{falls } w = \mathtt{11} \cdot w', \\ \varepsilon, & \text{falls } w = \varepsilon, \\ \perp, & \text{sonst} \end{cases}$$

d)

$$l_g\colon B^* \to (A \cup \{\mathtt{\perp}\})^*$$
$$w \mapsto \begin{cases} l_g(w') \cdot \mathtt{a}, & \text{falls } w = w' \cdot \mathtt{00} \\ l_g(w') \cdot \mathtt{b}, & \text{falls } w = w' \cdot \mathtt{10} \\ l_g(w') \cdot \mathtt{c}, & \text{falls } w = w' \cdot \mathtt{01} \\ l_g(w') \cdot \mathtt{d}, & \text{falls } w = w' \cdot \mathtt{011}, \\ \varepsilon, & \text{falls } w = \varepsilon, \\ \perp, & \text{sonst} \end{cases}$$

e)

x	a	b	c	d
$h(x)$	00	10	01	01

Der durch h induzierte Homomorphismus h^{**} kann nun keine Decodierung besitzen, denn h ist nicht injektiv und also keine Codierung. Es ist $h^{**}(\mathtt{c}) = \mathtt{01} = h^* * (\mathtt{d})$, also ist es für das Wort $\mathtt{01}$ nicht möglich, das Element aus A zu bestimmen, das von h^{**} darauf abgebildet wurde.

Aufgabe 2 - Blockschrift (5 Punkte)

In Kapitel 8 der Vorlesung wurde erwähnt, dass es eine bijektive Übersetzung zwischen Wörtern der Länge 4 in binärer Zahlenrepräsentation und Wörtern der Länge 1 in hexadezimaler Repräsentation gibt. In dieser Aufgabe wollen wir weitere solche Übersetzungen finden.

Sei dazu $k \in \mathbb{N}_+$ mit $k \geq 2$. Ein Wort der Länge $n \in \mathbb{N}_+$ über dem Alphabet Z_k wird in dieser Aufgabe als k-ärer n-Block bezeichnet.

a) In dieser Teilaufgabe sei zunächst $k = 16, l = 8$. Geben Sie die folgenden Elemente von Z_{16}^3, Z_8^4 an: (1 Punkt)

- $x_1 \in Z_{16}^3$ mit $\mathrm{Num}_{16}(x_1) = 10$
- $x_2 \in Z_8^3$ mit $\mathrm{Num}_8(x_2) = 10$
- $x_3 \in Z_{16}^4$ mit $\mathrm{Num}_{16}(x_3) = 4087$
- $x_4 \in Z_8^4$ mit $\mathrm{Num}_8(x_4) = 4087$

b) Gibt es eine Funktion $s : Z_{16}^3 \to Z_8^4$, die (1) bijektiv und (2) eine Übersetzung ist? Begründen Sie Ihre Antwort. (1 Punkt)

c) Seien $k, l, m, n \in \mathbb{N}_+$ mit $k \geq 2$ und $l \geq 2$. Geben Sie eine Bedingung an, die notwendig und hinreichend dafür ist, dass eine bijektive Übersetzung $t : Z_k^n \to Z_l^m$ existiert. (1 Punkt)

d) Zeigen Sie, dass Ihre Bedingung aus der vorherigen Teilaufgabe hinreichend ist. (2 Punkte)

Lösung 2

a) $x_1 = \texttt{00A}$ $\quad x_2 = \texttt{012}$ $\quad x_3 = \texttt{0FF7}$ $\quad x_4 = \texttt{7767}$

b) Ja, denn: Sei die Funktion t wie folgt definiert:

$$\begin{aligned} t\colon Z_{16}^3 &\to Z_8^4 \\ w &\mapsto \begin{cases} \mathrm{Repr}_8(\mathrm{Num}_{16}(w)), & \text{falls } |\mathrm{Repr}_8(\mathrm{Num}_{16}(w))| = 4 \\ \texttt{0} \cdot \mathrm{Repr}_8(\mathrm{Num}_{16}(w)), & \text{falls } |\mathrm{Repr}_8(\mathrm{Num}_{16}(w))| = 3 \\ \texttt{00} \cdot \mathrm{Repr}_8(\mathrm{Num}_{16}(w)), & \text{falls } |\mathrm{Repr}_8(\mathrm{Num}_{16}(w))| = 2 \\ \texttt{000} \cdot \mathrm{Repr}_8(\mathrm{Num}_{16}(w)), & \text{falls } |\mathrm{Repr}_8(\mathrm{Num}_{16}(w))| = 1 \end{cases} \\ &= \texttt{0}^{4-|\mathrm{Repr}_8(\mathrm{Num}_{16})(w)|} \cdot \mathrm{Repr}_8(\mathrm{Num}_8)(w) \end{aligned}$$

Dann ist t eine Übersetzung von Z_{16}^3 nach Z_8^4, die die Bedeutungen, in diesem Fall die Werte aus $\mathbb{Z}_{4096}$, der Wörter erhält.

Hierbei ist zu beachten, dass $|Z_8^4| = |Z_{16}^3|$, denn

$$\max_{w \in Z_{16}^3} \{\mathrm{Num}_{16}(w)\} = \mathrm{Num}_{16}(\texttt{FFF}) = 4095 = \mathrm{Num}_8(\texttt{7777}) = \max_{w \in Z_8^4} \{\mathrm{Num}_8(w)\}$$

Damit ist t bijektiv.

c) Es muss gelten, dass $|Z_k^n| = |Z_l^m|$, also $\max_{w \in Z_k^n} \{\mathrm{Num}_k(w)\} = \max_{w \in Z_l^m} \{\mathrm{Num}_l(w)\}$. Eine notwendige und hinreichende Bedingung ist somit, dass $k^n = l^m$ gelten muss.

d) Um zu zeigen, dass für $k^n = l^m$ (*) es die gesuchte bijektive Übersetzung gibt, betrachten wir die Funktion

$$\begin{aligned} t_{k,l,n,m} :& Z_k^n \to Z_l^m \\ & w \mapsto \mathtt{0}^{m-|\mathrm{Trans}_{l,k}(w)|} \cdot \mathrm{Trans}_{l,k}(w) \quad , \end{aligned}$$

die eine solche bijektive Übersetzung ist.

Sei $n = \mathrm{Num}_k(w)$. Es gilt $\mathrm{Trans}_{l,k}(w) = \mathrm{Repr}_l(Num_k(w)) = \mathrm{Repr}_l(n)$. Wegen $w \in Z_k^n$, ist $0 \leq n < k^n \overset{(*)}{=} l^m$. Deswegen ist $1 \leq |\mathrm{Repr}_l(n)| < m$. Damit ist die Definition von $t_{k,l,n,m}$ wohldefiniert, da $\mathtt{0}^{m-|\mathrm{Trans}_{l,k}(w)|} \cdot \mathrm{Trans}_{l,k}(w) \in Z_l^m$.

Man kann sich leicht überzeugen, dass die Funktion surjektiv ist, und da Definitions- und Zielbereich dieselbe Kardinalität haben, ist sie damit auch bijektiv.

Aus der Vorlesung ist bekannt, dass jede Trans-Funktion eine Übersetzung ist.

Zur Notwendigkeit (nicht gefordert): Es gilt zu zeigen, dass, wenn eine bijektive Übersetzung zwischen Z_k^n und Z_l^m existiert, $k^n = l^m$ gelten muss.

Sei also t eine bijektive Übersetzung zwischen Z_k^n und Z_l^m, das bedeutet insbesondere, dass $|Z_k^n| = |Z_l^m|$. Daraus folgt direkt, dass

$$k^n = |Z_k|^n = |Z_k^n| = |Z_l^m| = |Z_l|^m = l^m$$

Aufgabe 3 - Auf heißer Spur (5 Punkte)

Detektivin Lovelace ist noch immer davon überzeugt, dass Agent Hoare kein Spion war und ist entschlossen, ihn von den üblen Anschuldigungen gegen ihn reinzuwaschen. Dass Agent Zuse nur schlampig gegen ihn ermittelt zu haben scheint, bestärkt sie nur in diesem Vorhaben.

Verborgen in einer geheimen Schublade seines Schreibtisches findet Lovelace einen an sie adressierten Brief. Hoare scheint ihr eine letzte Nachricht hinterlassen zu haben! Doch ihre Enttäuschung ist groß, als der Umschlag nur einen Zettel mit wenigen Zeilen enthält, alle davon offenbar Kauderwelsch. Zuoberst steht das Wort

`verschluesslemich`

a) Stellen Sie einen Huffman-Baum zu diesem Wort auf und geben Sie die daraus resultierende Huffman-Codierung jedes auftrenden Zeichens an. (2 Punkte)

Die Detektivin ist enttäuscht, dass diese Spur nicht zu einem Ziel zu führen scheint. Auf dem Zettel steht aber auch noch das Folgende:

```
0001111110100100000010110
11101110100001001100
100010110001011
```

Was wollte der verstorbene Agent ihr wohl damit mitteilen? Helfen Sie der Detektivin, das herauszufinden!

Auf der Innenseite des Briefumschlags findet Lovelace etwas, das wie die unvollständige Definition einer Codierung aussieht:

x	a	e	h	n	r	s	t	u	v	z
N_x	2	3	1	2	2	2			1	
$h_m(x)$			00000	011	100	101	110	111		0001

„Aha! Hoare wollte uns einen Hinweis geben, wie wir seinen Code entschlüsseln können!", meint sie.

b) Sei $c \in Z_2^*$ die oben stehende Code-Nachricht. Die partielle Tabelle gibt die absoluten Häufigkeiten N_x der Buchstaben der Ursprungsnachricht $m \in \{\texttt{a}, \texttt{e}, \texttt{h}, \texttt{n}, \texttt{r}, \texttt{s}, \texttt{t}, \texttt{u}, \texttt{v}, \texttt{z}\}^*$ an. Die Funktion h_m ist eine Huffman-Codierung für m.

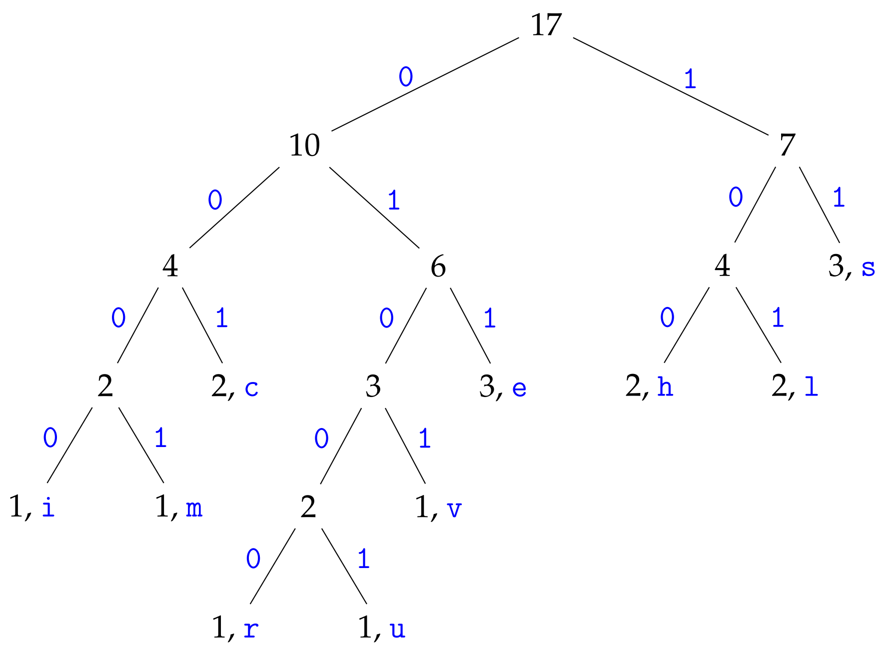
Rekonstruieren Sie die fehlenden Einträge in der Tabelle. Sie sind unter der Annahme, dass h_m eine Huffman-Codierung ist, eindeutig.

Geben Sie das Urbild m von c unter h_m mit $h_m(m) = c$ an. (3 Punkte)

Zufrieden lehnt sich Lovelace in ihrem Schreibtischstuhl zurück: „Na also, geht doch."

Lösung 3

a) Ein möglicher Huffman-Baum zu w ist:



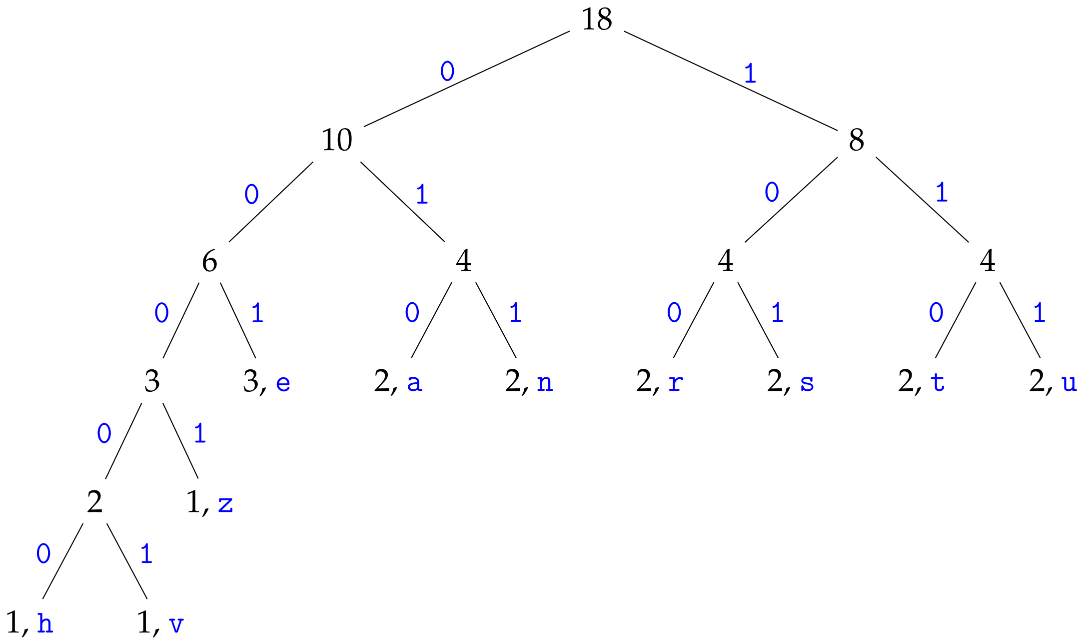
Daraus ergibt sich die folgende Huffman-Codierung h_w:

x	c	e	h	i	l	m	r	s	u	v
N_x	2	3	2	1	2	1	1	3	1	1
$h_w(x)$	001	011	100	0000	101	0001	01000	11	01001	0101

b) Wir vervollständigen zunächst die in der Aufgabenstellung angegebene Tabelle:

x	a	e	h	n	r	s	t	u	v	z
N_x	2	3	1	2	2	2	2	2	1	1
$h_m(x)$	010	001	00000	011	100	101	110	111	00001	0001

Der zugehörige Huffman-Baum ist:



Also ist c = zusehatunsverraten

Aufgabe 4 - Der Weisheit letzter Schluss (5 Punkte)

In der Vorlesung wurden bisher Huffman-Codierungen betrachtet, die auf ein bestimmtes Wort zugeschnitten sind. Doch was passiert, wenn man eine Codierung festlegen möchte, um damit bisher unbekannte Texte zu codieren?

In dieser Aufgabe sei das deutsche Alphabet $A = \{\mathtt{a}, \mathtt{b}, \ldots, \mathtt{z}\}$ der Buchstaben ohne Umlaute und ß gegeben. Es gilt also $|A| = 26$.

a) Durch einen Homomorphismus $h_{fix} : A^* \to Z_2^*$ soll jeder Buchstabe auf ein Wort derselben Länge $n \in \mathbb{N}_+$ abgebildet werden, so dass $|h_{fix}(x)| = n$ für alle $x \in A$. Welchen Wert muss n mindestens annehmen, damit die Übersetzung h_{fix} ein Code ist? Begründen Sie Ihre Antwort kurz. (1 Punkt)

b) Informieren Sie sich über die relative Häufigkeiten der Zeichen aus A, zum Beispiel unter Wikipedia[1]. Erstellen Sie darauf basierend eine Huffman-Codierung $h_{huffman} : A^* \to Z_2^*$ für deutsche Texte. Vervollständigen Sie dazu den Huffman-Baum in Abbildung 1, indem Sie die Teilbäume $*_1$ und $*_2$ angeben, und behandeln Sie die relativen Zeichenhäufigkeiten so wie die Anzahlen der Zeichenvorkommmen in einem konkreten Wort. (2 Punkte)

[1] Dort ist auch ß aufgelistet. Aber die Häufigkeit der Buchstaben ohne ß addiert sich schon auf 100%

c) Wie viele Zeichen aus Z_2 hat $h_{huffman}(\mathfrak{T})$ für einen deutschen Text $\mathfrak{T}$ mit $|\mathfrak{T}| = n$, der der angegebenen Buchstabenverteilung folgt, im Durchschnitt? (1 Punkt)

d) Betrachten Sie den folgenden Text $\mathfrak{F}$ aus *Faust. Der Tragödie zweiter Teil* von Johann Wolfgang von Goethe:

Daran erkenn' ich den gelehrten Herrn!
Was ihr nicht tastet, steht euch meilenfern,
Was ihr nicht fasst, das fehlt euch ganz und gar,
Was ihr nicht rechnet, glaubt ihr, sei nicht wahr,
Was ihr nicht waegt, hat fuer euch kein Gewicht,
Was ihr nicht muenzt, das, meint ihr, gelte nicht.

Die Anzahlen der Zeichenvorkommmen für diesen Text finden Sie in Tabelle 1.

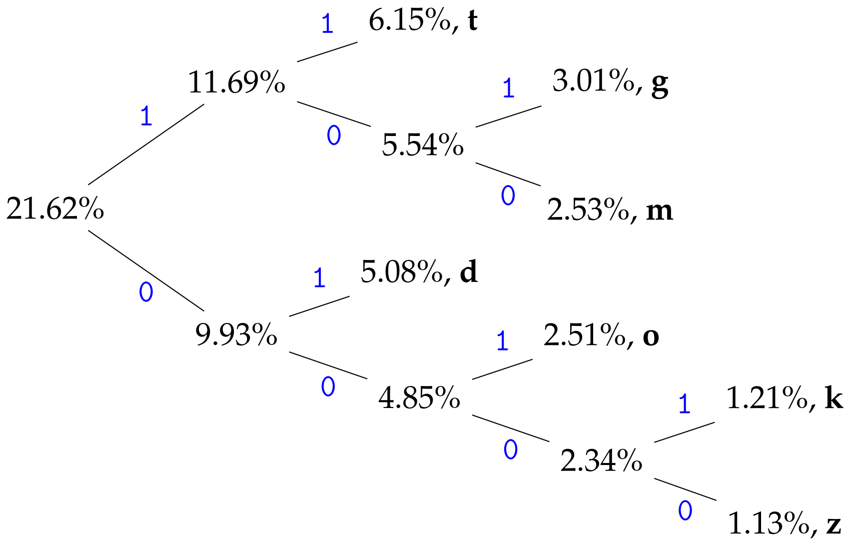
Geben Sie jeweils die Längen der Buchstaben-Codierung $h_{fix}(\mathfrak{F})$ aus Teilaufgabe (a) und der Huffman-Codierung $h_{huffman}(\mathfrak{F})$ aus Teilaufgabe (b) an. (1 Punkt)

Lösung 4

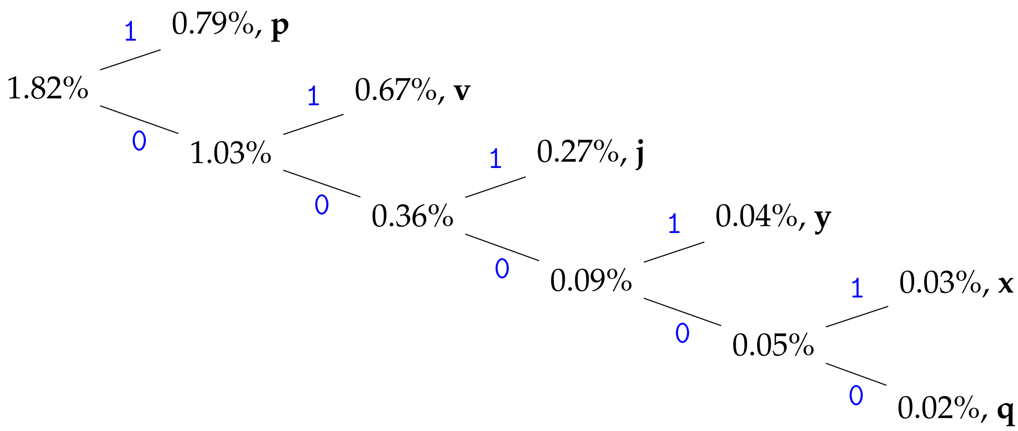
a) Mit $n \in \mathbb{N}_+$ Bits lassen sich 2^n verschiedene Werte darstellen. Um 26 Zeichen codieren zu können, muss n also so gewählt werden, das $2^n \geq 26$ ist. Der kleinstmögliche Wert für n ist damit 5, denn $2^4 = 16 < 26 < 32 = 2^5$.

b) Der Huffman-Baum aus Abbildung 1 lässt sich vervollständigen durch die folgenden beiden Teilbäume:

Der Teilbaum $*_1$ =



Der Teilbaum $*_2$ =



c) Mit der Lösung für Teilaufgabe (b) ergeben sich Codes der folgenden Längen:

x	a	b	c	d	e	f	g	h	i	j	k	l	m
$\lvert h_{huffman}(x) \rvert$	4	6	5	4	3	6	5	4	4	9	6	5	5

x	n	o	p	q	r	s	t	u	v	w	x	y	z
$\lvert h_{huffman}(x) \rvert$	3	5	7	11	4	4	4	4	8	6	8	7	6

Die durchschnittliche Länge der Codierung eines Textes mit der angege-

x	a	b	c	d	e	f	g	h	i	j	k	l	m
N_x	17	1	13	5	27	4	7	26	20	0	2	5	3

x	n	o	p	q	r	s	t	u	v	w	x	y	z
N_x	21	0	0	0	17	12	23	7	0	8	0	0	2

Tabelle 1: Die Anzahlen der Vorkommen der Zeichen aus A im gegebenen Beispieltext

benen Buchstabenverteilung ist

$$
\begin{aligned}
n \cdot \sum_{x \in A} N_x \cdot |h(x)| = 3n(N_e + N_n) \\
&+ 4n(N_a + N_d + N_h + N_i + N_r + N_s + N_t + N_u) \\
&+ 5n(N_c + N_g + N_l + N_m + N_o) \\
&+ 6n(N_b + N_f + N_k + N_w + N_z) \\
&+ 7nN_p + 8nN_v + 9nN_j + 10nN_y \\
&+ 11n(N_q + N_x)
\end{aligned}
$$

Dabei ist N_x hier die relativen Häufigkeit von x.

Damit ergibt sich eine durchschnittliche Codelänge von $n \cdot 4.0992$.

d) Der gesamte Text umfasst 220 Zeichen. Also ist nach dem Ergebnis aus Teilaufgabe (a) $|h_{fix}(\mathfrak{F})| = 5 \cdot 220 = 1100$.

Die Länge der Huffman-Codierung des Textes mit der Codierung aus Teilaufgabe (b) hätte die folgende Länge:

$$
\begin{aligned}
\sum_{x \in A} N_x \cdot |h(x)| = 3(N_e + N_n) \\
&+ 4(N_a + N_d + N_h + N_i + N_r + N_s + N_t + N_u) \\
&+ 5(N_c + N_g + N_l + N_m + N_o) \\
&+ 6(N_b + N_f + N_k + N_w + N_z) \\
&+ 7N_p + 8N_v + 9N_j + 10N_y \\
&+ 11(N_q + N_x) \\
&= 3 \cdot 48 + 4 \cdot 127 + 5 \cdot 28 + 6 \cdot 17 \\
&= 894
\end{aligned}
$$

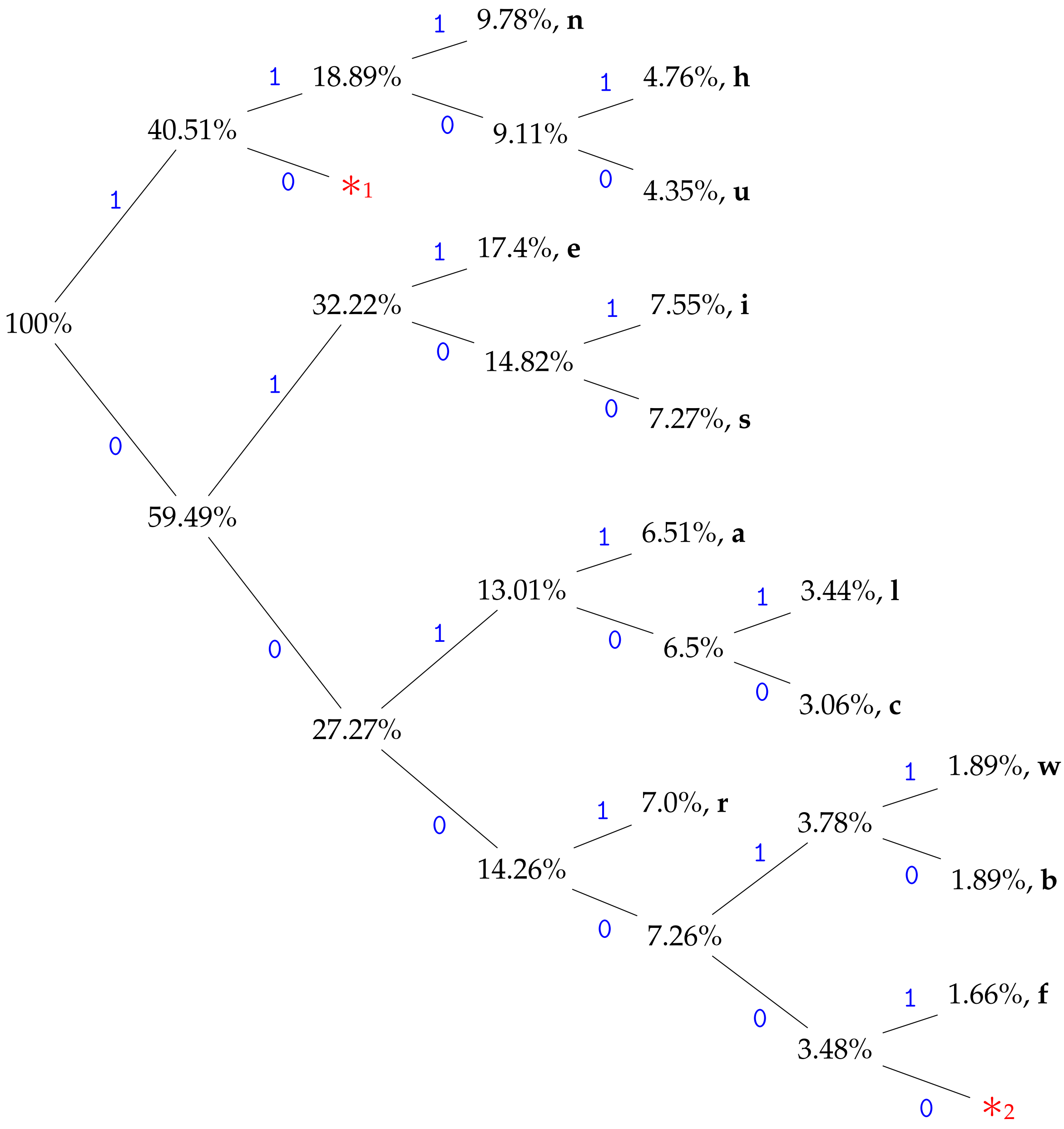


Abbildung 1: Unvollständiger Huffman-Baum zur relativen Zeichenhäufigkeit in deutschen Texten

Bitte beachten Sie die folgenden Hinweise:

- Lösungen **müssen** handschriftlich erstellt werden
- Ihre Abgabe sollte die erste Seite dieser Datei als Deckblatt haben
- Ihre Abgabe muss **rechtzeitig** erfolgen

Außerdem, wenn Sie Ihre Ausarbeitung über die Abgabekästen im Keller des Informatik-Gebäudes abgeben:

- Ihre Abgabe muss in der oberen **linken** Ecke zusammengeheftet werden
- Tablet-Ausdrucke sind zulässig

Wenn Sie Ihre Ausarbeitung online über ILIAS abgeben, dann achten Sie darauf:

- Ihre Abgabe muss **genau eine** PDF-Datei sein
- Scans und lesbare Fotos sind zulässig
- Abgabe erfolgt unter „Tutorien" im Ordner **Ihres** Tutoriums