1 Grundbegriffe der Codierungstheorie

Es sei $A = \{a_1, a_2, \ldots, a_q\}$, $q \in \mathbb{N}, q \geq 2$ eine endliche Menge mit q Elementen. Wir nennen A einen **q-ären Zeichenvorrat** oder ein **Alphabet**. Die Elemente aus A heißen **Zeichen** oder **Codebuchstaben**.

Beispiel 1.1. Beispiele für Alphabete:

– $\{\mathrm{a}, \mathrm{b}, \mathrm{c}, \ldots, \mathrm{y}, \mathrm{z}\}$	gewöhnliches Alphabet
– $\{0, 1, 2, \ldots, 8, 9\}$	Dezimalziffern
– $\{0, 1, 2, \ldots, 8, 9, A, B, C, D, E, F\}$	Hexadezimalziffern
– $\{\cdot, -, \text{Pause}\}$	Zeichen des Morse-Codes
– $\{A, G, C, T\} \cup \{\emptyset\}$	Zeichen für die Nukleotiden (Adenin, Guanin, Cytosin, Thymin) des genetischen Codes im DNS-Molekül einschließlich einer „dummy-Variablen“ für Auslassungen
– Zeichensatz des ASCII-Codes	

In der Nachrichtenübertragung werden als Codebuchstaben solche Zeichen verwendet, die der Kanal übertragen kann. In der Technik ist es besonders einfach, nur zwei physikalisch unterscheidbare Zustände darzustellen.

Im Fall $q = 2$ nennt man A einen **binären Zeichenvorrat** oder ein **Binäralphabet**. Die Zeichen eines binären Zeichenvorrats heißen **Binärzeichen** oder **Bits**.

Beispiel 1.2. Beispiele für Binäralphabete:

- $A = \{\text{falsch}, \text{wahr}\}$
- $A = \{\text{„Strom fließt“}, \text{„kein Strom fließt“}\}$
- $A = \{0, 1\}$

Sei $n \in \mathbb{N}$ eine natürliche Zahl und A ein Alphabet. Unter einem **Wort der Länge n über dem Alphabet A** versteht man ein n-Tupel $z = (z_1, z_2, \ldots, z_n)$ von Elementen aus A. Es wird meist ohne Klammern und Beistriche geschrieben: $z = z_1 z_2 \ldots z_n$.
(Dabei heißt z_i die i-te Komponente des Wortes.)

Beispiel 1.3.

– $A = \{a, b, c, \ldots, y, z\}$	*treffen* ist ein Wort der Länge 7 über A
– $M = \{\cdot, -, \text{Pause}\}$	$\cdot\cdot\cdot$Pause $- - -$ Pause$\cdot\cdot\cdot$Pause ist ein Wort der Länge 9 über dem Morsealphabet M (und bedeutet „SOS“)

1.1 Nachrichtenquelle

Beispiel 1.4. Hier einige Beispiele für verschiedene Nachrichtenquellen:

a) ein Autor, der ein Buch schreibt

b) ein Klavierspieler, der ein Stück spielt

c) eine Person, die mit jemandem am Telefon spricht

Eine Nachrichtenquelle Q produziert laufend eines der folgenden Wörter: $x_1, \ldots, x_n$ ($n \in \mathbb{N}$). Die Menge $N = \{x_1, x_2, \ldots, x_n\}$ bezeichnet man als **Nachrichtenvorrat**.

Ein Element aus der Menge N nennt man **Nachricht**.

Der Nachrichtenvorrat kann verschiedene Formen annehmen.

Beispiel 1.5. Alice und Bob schreiben einander SMS.

- Nachrichtenquelle: der jeweilige Schreiber (d. h. Alice und Bob) wechseln sich in der Rolle der Nachrichtenquelle ab
- Nachrichtenvorrat: $A = \{0, 1, 2, \ldots, 9$, a, b, c, $\ldots$, z, „Leerzeichen“, ., !, ?, $\ldots\}$ (alphanumerischer Zeichenvorrat inklusive Satz- und Sonderzeichen) ist das Alphabet der Quelle.
  Nachrichtenvorrat $N = A^1 = A$ (alle Wörter der Länge 1 über dem Alphabet A)
- Nachrichtensequenz: Eine Nachrichtensequenz ist eine Sequenz von Elementen aus der Menge N, z. B.: „hallo“, „wie geht es dir“, …

1.1.1 Modellierung einer Nachrichtenquelle

In der Informationstheorie erfolgt eine mathematische Beschreibung einer Nachrichtenquelle mit Hilfe der Wahrscheinlichkeitsrechnung.

Das Senden einer Nachricht kann als Zufallsexperiment gesehen werden. Es wird dabei aus dem Nachrichtenvorrat der Quelle jede Nachricht mit einer gewissen Wahrscheinlichkeit ausgewählt, d. h. die Nachrichtenquelle Q gibt jede der Nachrichten $x_1, x_2, \ldots, x_n$ mit einer Wahrscheinlichkeit von $p_1, p_2, \ldots, p_n$ aus.

Meist sendet eine Quelle keine einzelne Nachricht, sondern mehrere Nachrichten hintereinander – das Zufallsexperiment wird mehrmals hintereinander ausgeführt.

Beispiel 1.6. Eine Nachrichtenquelle verfügt über den Nachrichtenvorrat $N = \{0, 10, 11\}$ und sendet pro Zeiteinheit eine Nachricht. Die Nachricht 0 wird mit der Wahrscheinlichkeit 0,5, die Nachricht 10 mit der Wahrscheinlichkeit 0,3 und die Nachricht 11 mit der Wahrscheinlichkeit 0,2 gesendet.

- $A_1 = \{0, 11\}$ ist das Ereignis, dass die Nachrichtenquelle entweder die Nachricht 0 oder die Nachricht 11 sendet. Für die Wahrscheinlichkeit dieses Ereignisses gilt:

$$p(A_1) = p(\{0\} \cup \{11\}) = p(\{0\}) + p(\{11\}) = 0{,}7$$

- Wir berechnen die Wahrscheinlichkeit dafür, dass die Nachrichtenquelle die Nachrichtensequenz $0 * 10 = 010$ sendet.
  Das Zufallsexperiment ist das zweimalige Senden einer Nachricht. Wir berechnen die Wahrscheinlichkeit des Ereignisses, dass zuerst die Nachricht 0 und anschließend die Nachricht 10 gesendet wird. Dieses Ereignis lautet folgendermaßen: $A_2 = \{(0, 10)\}$. Da die hintereinander ausgeführten Teilexperimente unabhängig sind, gilt für die Wahrscheinlichkeit des Ereignisses A_2:

$$p(A_2) = p_1(\{0\}) \cdot p_2(\{10\}) = 0{,}5 \cdot 0{,}3 = 0{,}15$$

Übung 1.1. Die fünfjährige Julia hat bisher folgende Buchstaben gelernt: A, B, D, K, M, O und R. Aus diesen formt sie verschieden lange „Wörter“. Sie wählt dabei die Buchstaben A, D, O mit der Wahrscheinlichkeit 0,2, die Buchstaben B und M mit der Wahrscheinlichkeit 0,15 und die Buchstaben K und R mit der Wahrscheinlichkeit 0,05 unabhängig voneinander aus. Wie groß ist die Wahrscheinlichkeit, dass Julia

a) das Wort ABRAKADABRA schreibt?
b) entweder OMA oder MAMA schreibt?

1.2 Codierung

Einige bekannte Beispiele für Codierungen:

Tonverarbeitung (CD – Compact Disc): Eine Tonquelle liefert ein analoges Signal. Auf einer Audio-CD sind die Daten jedoch in digitaler Form gespeichert. Dazu wird das analoge Signal der Quelle mittels Diskretisierung (Abtastung und Quantisierung) in ein digitales Signal umgewandelt (Analog-Digital-Wandlung). Es entsteht eine Folge von Binärzeichen (Bits), die auf der CD gespeichert wird.

ASCII-Code: Der ASCII-Code ist der übliche Standard zur binären Darstellung von Buchstaben, Zahlen und Zeichen in einem Computer.

Morse-Code: Im Morse-Code werden den einzelnen Zeichen des alphanumerischen Zeichenvorrats Wörter, die aus den Zeichen „·“ und „–“ bestehen, zugeordnet. Zusätzlich existiert noch ein Zeichen, die „Pause“, die wir hier als Komma „ ,“ darstellen. Die „Pause“ („ ,“) dient dazu, um das Ende eines Zeichens des alphanumerischen Zeichenvorrats zu markieren. Korrekterweise müssten wir z. B. das Morse-Codewort für den Buchstaben „y“ als „– · – – ,“ und für das Leerzeichen als „ , , ,“ angegeben.

Tabelle 1.1 Auszug aus der Codetabelle des Morse-Codes

a	· –	m	– –	x	– · · –	.	· – · – · –
ä	· – · –	n	– ·	y	– · – –	,	– – · · – –
b	– · · ·	o	– – –	z	– – · ·	:	– – – · · ·
c	– · – ·	ö	– – – ·	0	– – – – –		
d	– · ·	p	· – – ·	1	· – – – –		
e	·	q	– – · –	2	· · – – –	Leerzeichen	, ,
f	· · – ·	r	· – ·	3	· · · – –	Anfang	– · – · –
g	– – ·	s	· · ·	4	· · · · –	Ende	· · · – · –
h	· · · ·	t	–	5	· · · · ·		
i	· ·	u	· · –	6	– · · · ·		
j	· – – –	ü	· · – –	7	– – · · ·		
k	– · –	v	· · · –	8	– – – · ·		
l	· – · ·	w	· – –	9	– – – – ·		

Aus technischen oder anderen Gründen ist es oft notwendig, vom ursprünglichen Alphabet der Quelle zu einem anderen Alphabet überzugehen. Dazu codiert man die Buchstaben eines Alphabets A durch Wörter über einem anderen Alphabet B.

In der Praxis besonders wichtig ist die Codierung durch Wörter über dem Alphabet $B = \{0, 1\}$. Daher beschränken wir uns auf diese.

Es sei A ein beliebiges und $B = \{0, 1\}$ ein binäres Alphabet und $n \in \mathbb{N}$. Eine Codierung des Alphabets A durch Wörter über B ist eine Funktion

$$c : A \longrightarrow B^1 \cup B^2 \cup \ldots B^{n-1} \cup B^n = \bigcup_{i=1}^{n} B^i,$$

bei der verschiedene Zeichen aus A auch durch verschiedene Wörter über B dargestellt werden. Die Elemente von $c(A)$ heißen **Codewörter**. Die Menge aller Codewörter ($= c(A)$) wird auch als **Code** bezeichnet.

Beispiel 1.7. Gegeben sind die Alphabete $A = \{a, b, c, d\}$ und $B = \{0, 1\}$. Wir definieren 4 verschiedene Codierungen $c_1, \ldots, c_4$ von A durch Wörter über B mit folgender Tabelle:

$x \in A$	I. $c_1(x)$	II. $c_2(x)$	III. $c_3(x)$	IV. $c_4(x)$
a	00	0	0	010
b	01	10	01	11000
c	10	110	10	01000
d	11	111	11	00110

$c_1(A) = \{00, 01, 10, 11\}$ ist ein **Blockcode der Länge** 2.
$c_2(A), c_3(A)$ sowie $c_4(A)$ sind **Codes variabler Länge**.

Ein Code heißt **Blockcode** (der Länge n), wenn alle seine Codewörter dieselbe Länge n haben. Ansonsten spricht man von einem **Code variabler Länge**. Zur Quellencodierung werden vor allem Codes variabler Länge eingesetzt, während für die Kanalcodierung Blockcodes bevorzugt werden.

Tabelle 1.2 Auszug aus einer ASCII-Code-Tabelle

0	0110000	@	1000000	P	1010000
1	0110001	A	1000001	Q	1010001
2	0110010	B	1000010	R	1010010
3	0110011	C	1000011	S	1010011
4	0110100	D	1000100	T	1010100
5	0110101	E	1000101	U	1010101
6	0110110	F	1000110	V	1010110
7	0110111	G	1000111	W	1010111
8	0111000	H	1001000	X	1011000
9	0111001	I	1001001	Y	1011001
:	0111010	J	1001010	Z	1011010
;	0111011	K	1001011	[	1011011
<	0111100	L	1001100	\	1011100
=	0111101	M	1001101	]	1011101
>	0111110	N	1001110	^	1011110
?	0111111	O	1001111	_	1011111

Bemerkung
Die heute übliche Form des ASCII-Codes („American Standard Code for Information Interchange“ – „Amerikanische Normdarstellung zum Informationsaustausch“) stammt aus dem Jahre 1968. Beim ASCII-Code wird jedem von 128 Zeichen ein Tupel aus sieben Bitwerten zugeordnet. Schon früh haben Computer oft Acht-Bit-Tupel verwendet – wobei das achte Bit für Fehlerkorrekturzwecke auf den Kommunikationsleitungen oder für andere Steuerungsaufgaben verwendet wurde.

ASCII diente ursprünglich zur Darstellung von Schriftzeichen der englischen Sprache. Um später auch Sonderzeichen anderer Sprachen darstellen zu können (z. B. deutsche Umlaute), wurde der Code von sieben auf acht Bit erweitert. Allerdings bot auch der Acht-Bit-Code zu wenig Platz, um alle Sonderzeichen gleichzeitig unterzubringen, wodurch mehrere verschiedene Erweiterungen notwendig wurden. Keine dieser Erweiterungen sollte aber als ASCII bezeichnet werden, um Verwirrung zu vermeiden – ASCII bezeichnet nur den Sieben-Bit-Code.)

HÜbung 1.1. Decodiere folgende, mit dem Morse-Code codierte Nachricht:
·−−, ··, ·−·,,,, −, ·− ·, ·, ··−·, ··−·, ·, −·,,,, ··−, −·, ···,,,, −−, −−−, ·−
·, −−·, ·, −·,,,, ··−, −−,,,, −−−··,,,, ··−, ····, ·−·, ·−·−·−,

2 Grundlagen der Informationstheorie

Der Begriff „Information“, wie er in der Informationstheorie gebraucht wird, unterscheidet sich grundsätzlich vom umgangssprachlichen Informationsbegriff. Die Informationstheorie liefert eine mathematisch-physikalische Definition der „Information“ und macht sie so zu einer messbaren Größe wie Spannung, Temperatur, usw.

Problemstellung
Eine Nachrichtenquelle Q mit dem Nachrichtenvorrat $N = \{x_1, x_2, \dots, x_n\}$ sendet pro Zeiteinheit eine Nachricht. Die Wahrscheinlichkeit, dass die i-te Nachricht x_i gesendet wird ist p_i. Welchen Informationsgehalt hat die i-te Nachricht?

Betrachten wir dazu ein konkretes Beispiel:

Beispiel 2.1. Beim Lotto „6 aus 45“ werden nacheinander 7 Zahlen (6 und die Zusatzzahl) gezogen. Die Lostrommel entspricht dabei einer Nachrichtenquelle mit dem Nachrichtenvorrat $N = \{1, 2, 3, \dots, 44, 45\}$. Die Wahrscheinlichkeit p_i, dass die Zahl i (z. B. 10) gezogen wird, beträgt für alle Zahlen $\frac{1}{45}$. Welchen Informationsgehalt haben die einzelnen Zahlen?

Eine Nachrichtenquelle Q mit dem Zeichenvorrat $N = \{x_1, x_2, \dots, x_n\}$ sendet pro Zeiteinheit eine Nachricht. Die Wahrscheinlichkeit, dass die i-te Nachricht x_i gesendet wird ist p_i. Der Informationsgehalt einer Nachricht x_i mit der Wahrscheinlichkeit p_i ist

$$\begin{aligned}
I :\quad &]0,1] \longrightarrow \mathbb{R} \\
I(p_i) &= {}^2\log\left(\frac{1}{p_i}\right) \\
&= -\,{}^2\log(p_i) = -\,\mathrm{ld}(p_i) \qquad ({}^2\log(x) = \mathrm{ld}(x) \dots \text{Logarithmus dualis})
\end{aligned}$$

Maßeinheit: $[I] = \text{bit}$ (kleingeschrieben)

Übung 2.1. Eine Nachrichtenquelle verfügt über den Nachrichtenvorrat $N = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ mit den in der Tabelle zugeordneten Wahrscheinlichkeiten. Berechne den Informationsgehalt der einzelnen Nachrichten.

Nachricht	x_1	x_2	x_3	x_4	x_5	x_6
Wahrscheinlichkeit	0,08	0,20	0,05	0,12	0,30	0,25

Tabelle 2.3 Buchstaben und deren Wahrscheinlichkeiten in der deutschen Schriftsprache

Buchstabe	Wahrscheinlichkeit	Buchstabe	Wahrscheinlichkeit
A	0,0651	N	0,0992
B	0,0257	O	0,0229
C	0,0284	P	0,0094
D	0,0541	Q	0,0007
E	0,1669	R	0,0654
F	0,0204	S	0,0678
G	0,0365	T	0,0674
H	0,0406	U	0,0370
I	0,0782	V	0,0107
J	0,0019	W	0,0140
K	0,0188	X	0,0002
L	0,0283	Y	0,0003
M	0,0301	Z	0,0100

Jetzt haben wir den Informationsgehalt einer Nachricht, die von einer Quelle ausgegeben wird, bestimmt. Im Allgemeinen ist man jedoch nicht am Informationsgehalt einer speziellen Nachricht interessiert, sondern am **Informationsgehalt einer Quelle**.

Gegeben sei eine Nachrichtenquelle Q mit dem Nachrichtenvorrat $N = \{x_1, x_2, \dots, x_n\}$ und den zugehörigen Wahrscheinlichkeiten $p_1, p_2, \dots, p_n$. Dann heißt

$$\begin{aligned} H(Q) &:= -p_1 \cdot {}^2\log(p_1) - p_2 \cdot {}^2\log(p_2) - \cdots - p_n \cdot {}^2\log(p_n) \\ &= -\sum_{i=1}^{n} p_i \cdot {}^2\log(p_i) \qquad [H] = \text{bit} \end{aligned}$$

die **Entropie der Quelle** Q.

Da die Entropie nur von den Wahrscheinlichkeiten abhängt, schreibt man auch $H(p_1, p_2, \dots, p_n)$.

Die Entropie misst unsere Unsicherheit, wenn wir raten sollen, welche Nachricht von der Quelle zu einem zukünftigen Zeitpunkt ausgewählt wird. Je größer die Entropie einer Quelle (d. h. die im Mittel pro Nachricht gelieferte Information) ist, desto größer ist unsere Unsicherheit.

Übung 2.2. Berechne $H(0{,}2; 0{,}2; 0{,}1; 0{,}5)$.

2.1 Interpretation der Entropie

Wie bereits erwähnt misst die Entropie unsere Unsicherheit, wenn wir raten sollen, welche Nachricht von der Nachrichtenquelle zu einem zukünftigen Zeitpunkt ausgewählt wird; dabei ist vorausgesetzt, dass wir die Wahrscheinlichkeiten der einzelnen Nachrichten kennen.

Wir haben die Entropie als Erwartungswert der Informationsgehalte aller Nachrichten definiert, d. h. die Entropie ist der mittlere Informationsgehalt pro Nachricht. Es gibt jedoch noch zwei weitere Möglichkeiten, die Entropie zu interpretieren:

1. Die Entropie gibt Antwort auf folgende Frage:
Wie viele Ja/Nein-Entscheidungen sind durchschnittlich *mindestens* notwendig, um die aktuelle Nachricht zu erfragen?
Zum Beispiel: Die Entropie der Binärquelle nimmt im Maximum den Wert 1 bit an. Dies entspricht genau einer Ja/Nein-Entscheidung, um die aktuelle Nachricht zu erfragen: Ist die Nachricht „0"?

2. Die Entropie gibt weiters Antwort auf folgende Frage:
Wie viele Bits (= Binärzeichen) benötigt man mindestens im Mittel, um die Nachrichten einer Quelle zu codieren?

3 Quellencodierung

Gegeben ist folgende Ausgangssituation:

1. Eine Nachrichtenquelle, die verschiedene Nachrichten produziert.
   Beispiele: Buchstaben des Alphabets, Ziffern, Wörter, ...

2. Ein Kanal, der in der Lage ist, Binärzeichen zu übertragen.

Dabei nehmen wir an, dass der Kanal störungsfrei ist (engl.: *noiseless channel*). Das bedeutet, dass bei der Übertragung von Nachrichten über den Kanal keine Fehler (Verzerrungen, Störungen) auftreten und wir daher kein Problem bezüglich Fehlererkennung und Fehlerkorrektur haben.

Ziel der Quellencodierung:

Die Nachrichten sollen nun in eine für die Übertragung geeignete, binäre Form gebracht werden. Dabei sollen folgende Forderungen beachtet werden:

Forderung 1: Die Codierung soll **effizient** sein!
Die Anzahl der Nachrichten, die in einer gegebenen Zeit über den Kanal übertragen werden, soll maximal sein.
Folgerung: Die Codewörter sollen so kurz wie möglich sein.

Forderung 2: Die Codierung soll **umkehrbar** sein!
Da die Datenübertragung selbst störungsfrei verläuft, können Fehler nur bei der Decodierung auftreten.
Folgerung: Aus der Folge der übertragenen Zeichen müssen die codierten Nachrichten eindeutig rekonstruiert werden können.

Zur Quellencodierung werden vor allem **Codes variabler Länge** eingesetzt.

Forderung 2 wird durch die Verwendung von sogenannten **Präfixcodes** erfüllt. Präfixcodes sind eindeutig decodierbar, d. h. jede endliche Folge von Binärzeichen (Bits) entspricht höchstens einer Nachrichtensequenz. Ein Präfixcode ist dadurch gekennzeichnet, dass kein Codewort Anfang eines anderen Codeworts ist.

Beispiel 3.1. Wir betrachten den Code $\mathcal{C} = \{0, 10, 110, 111\}$.

0	ist kein Präfix von	10, 110 und 111,	
10	ist kein Präfix von	0, 110 und 111,	
110	ist kein Präfix von	0, 10 und 111	und
111	ist kein Präfix von	0, 10 und 110.	

Somit ist $\mathcal{C}$ ein Präfixcode, da keines der Codewörter Präfix eines anderen Codeworts ist.

Übung 3.1. Ist der Code $\mathcal{C} = \{1, 00, 010, 0110, 0111\}$ ein Präfixcode? Begründe deine Antwort.

Durch die Verwendung von Präfixcodes ist Forderung 2 erfüllt. Eine weiteres Ziel der Quellencodierung ist eine möglichst effiziente Codierung, d. h. die durchschnittliche Codewortlänge eines Codes soll minimal sein. Es sei $k \in \mathbb{N}$ die Anzahl der Codewörter von einem Präfixcode $\mathcal{C} = \{c_1, c_2, \ldots, c_k\}$. Die Längen der einzelnen Codewörter sind $n_1, n_2, \ldots, n_k \in \mathbb{N}$ und $p_1, p_2, \ldots, p_k \in]0, 1]$ sind die Wahrscheinlichkeiten, mit denen die Codewörter ausgewählt werden. Dabei muss gelten: $\sum_{i=1}^{k} p_i = 1$. Die **durchschnittliche Wortlänge** $\overline{n}$ **von** $\mathcal{C}$ ist gegeben durch

$$\overline{n} = p_1 \cdot n_1 + p_2 \cdot n_2 + \cdots + p_k \cdot n_k$$

Übung 3.2. Eine Nachrichtenquelle Q sendet jedes der Codewörter $c_i \in \mathcal{C}$ mit der angegebenen Wahrscheinlichkeit aus. Berechne jeweils die durchschnittliche Wortlänge folgender Codes:

a) $\mathcal{C} = \{1, 00, 01\}$; Alle Codewörter sind gleich wahrscheinlich.

b)

Codewort $c_i \in \mathcal{C}$	11111	0	101	110
Wahrscheinlichkeit	0,05	0,7	0,15	0,1

c)

Codewort $c_i \in \mathcal{C}$	1	01111	01110	0110	010	00
Wahrscheinlichkeit	0,55	0,02	0,03	0,1	0,1	0,2

Es gilt:

$$\overline{n} \geq H(p_1, p_2, \ldots, p_k)$$

Ein Verfahren, mit dem man einen Präfixcode mit kleinster durchschnittlicher Wortlänge (sog. **optimaler Code**)konstruieren kann, ist die **Huffman-Codierung**.

3.1 Codierung nach Huffman

Beispiel 3.2. Gegeben: Nachrichtenvorrat $N = \{a, b, c, d, e, f\}$ und Binäralphabet $B = \{0, 1\}$;
Gesucht: Eine Codierung der Nachrichten aus N durch Wörter über B, die so effizient wie möglich ist.
Jede der Nachrichten $x_i \in N$ wird mit der Wahrscheinlichkeit p_i von der Nachrichtenquelle ausgegeben.

Nachricht x_i	a	b	c	d	e	f
Wahrscheinlichkeit p_i	0,05	0,15	0,05	0,4	0,15	0,2
Codewort						
Codewortlänge n_i						

Wie findet man eine solche Codierung?

Die Huffman-Codierung erfolgt in drei Schritten.

1. **Ordnen**: Ordne die Nachrichten nach fallenden Wahrscheinlichkeiten.
2. **Reduzieren**: Kombiniere* die beiden Nachrichten mit den kleinsten Wahrscheinlichkeiten zu einer neuen „zusammengesetzten Nachricht“. Ordne die Liste neu wie in Schritt 1 und fahre fort, bis alle Nachrichten zusammengefasst sind.
3. **Codieren**: Beginne bei der letzten Zusammenfassung; ordne der ersten Komponente der „zusammengefassten Nachricht“ als erste Ziffer des Codeworts eine „0“, der zweiten Komponente eine „1“ zu.

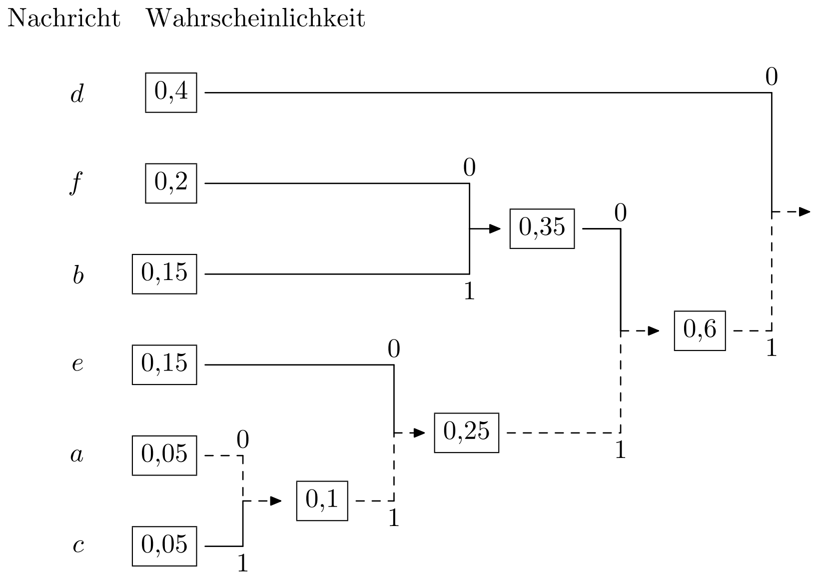
   Fahre sinngemäß fort, bis alle Nachrichten codiert sind.

*Im Falle mehrerer Nachrichten mit derselben Wahrscheinlichkeit werden jene Nachrichten kombiniert, die am wenigsten bereits zusammengefasste Nachrichten beinhalten. Damit erreicht man bei gleicher mittlerer Codewortlänge eine in der Übertragungstechnik günstigere, weil gleichmäßigere Verteilung der einzelnen Codewortlängen.

Lösung zu Beispiel 3.2.
Im ersten Schritt ordnen wir die Nachrichten nach fallenden Wahrscheinlichkeiten. Im zweiten Schritt werden die beiden Nachrichten mit den kleinsten Wahrscheinlichkeiten „c“ und „a“ kombiniert. Die beiden neuen Nachrichten mit den kleinsten Wahrscheinlichkeiten, „ca“ und „e“ werden jetzt zusammengefasst. Für die zusammengesetzte Nachricht erhält man die Wahrscheinlichkeit 0,25. Damit ist sie größer als die Wahrscheinlichkeiten für „b“ und „f“. Letztere sind nun die beiden kleinsten Wahrscheinlichkeiten. Es werden „b“ und „f“ zusammengefasst. Die zugehörige Wahrscheinlichkeit hat den Wert 0,35. Die nunmehr beiden kleinsten Wahrscheinlichkeiten, für „bf“ und „cae“, ergeben zusammen die Wahrscheinlichkeit 0,6. Die beiden verbleibenden Wahrscheinlichkeiten für „d“ und „$caebf$“ müssen zusammen den Wert eins ergeben.

Im dritten Schritt werden den Nachrichten die Codewörter zugewiesen. Hierbei beginnt man ganz rechts und schreitet nach links fort. Bei jeder Weggabelung (Zusammenfassung von Nachrichten) wird dem Pfad nach oben die „0“ und dem Pfad nach unten die „1“ (oder jeweils umgekehrt) zugewiesen. Der Pfad für die Codezuteilung für die Nachricht „a“ ist in Abbildung 3.1 strichliert gedruckt.

Abbildung 3.1 Huffman-Codierung



Man erhält schließlich folgenden Huffman-Code:

Nachricht x_i	a	b	c	d	e	f
Wahrscheinlichkeit p_i	0,05	0,15	0,05	0,4	0,15	0,2
Codewort	1110	101	1111	0	110	100
Codewortlänge n_i	4	3	4	1	3	3

Anwendung findet die Huffman-Codierung bei der Kompression von Texten, im Bilddaten-Kompressionsverfahren JPEG oder in der MP3-Codierung. Eine Datenreduzierung um 50 % ist dabei keine Seltenheit.

Übung 3.3. Eine Nachrichtenquelle Q verfügt über den Nachrichtenvorrat $N = \{x_1, \ldots, x_{13}\}$ mit den in der Tabelle zugeordneten Wahrscheinlichkeiten $p_1, \ldots, p_{13}$.

Nachricht x_i	x_1	x_2	x_3	x_4	x_5	x_6	x_7
Wahrscheinlichkeit p_i	0,2	0,18	0,1	0,1	0,1	0,061	0,059
Nachricht x_i	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	
Wahrscheinlichkeit p_i	0,04	0,04	0,04	0,04	0,03	0,01	

a) Führe eine Huffman-Codierung durch.
b) Berechne die Entropie der Quelle und die durchschnittliche Codewortlänge.