

Die Kompressionsrate ist so definiert, dass ein grösserer Wert eine bessere Kompression bedeutet.

Aufgabe 1

Die *Redundanz* beschreibt das Vorhandensein von Daten in einer Informationsquelle, die ohne Informationsverlust weggelassen werden können.

Beispiel: Natürliche Sprachen enthalten Redundanz, da wir natürlichsprachige Texte auch dann verstehen können, wenn Zeichen weggelassen oder verändert werden. Wir sollten also imstande sein, den folgenden Satz zu rekonstruieren, in dem es um Redundanz geht.

„Rdnndnt is dr Til enr Nchricht, dr kne Infrmtin nthlt.“

Aufgabe 2

Datenkompression ist ein Vorgang, bei dem die Menge digitaler Daten reduziert wird. Dadurch sinkt der Speicherbedarf und die Übertragungszeit.

Bei der Datenkompression wird versucht, redundante Informationen zu entfernen (*komprimieren*, *compress*), um sie später wieder hinzuzufügen (*dekomprimieren*, *decompress*).

Es werden zwei Arten von Kompression unterschieden:

- Bei der *verlustfreie Kompression* können die Originaldaten aus den komprimierten Daten exakt zurückgewonnen werden.

$$\text{decompress}(\text{compress}(\text{nachricht})) = \text{nachricht}$$

- Bei der *verlustbehaftete Kompression* können die Originaldaten aus den komprimierten Daten meist nicht exakt zurückgewonnen werden.

$$\text{decompress}(\text{compress}(\text{nachricht})) \approx \text{nachricht}$$

Aufgabe 3

Verlustbehaftete Kompression eignet sich hauptsächlich für Audio- und Videodaten, da die Originaldaten meist Informationen enthalten, welche vom Grossteil der Menschen nicht wahrgenommen werden können, wie z. B. sehr hohe Frequenzen. Somit wird eine unvollständige Rekonstruktion der ursprünglichen Daten nicht bemerkt.

Verlustbehaftete Kompression eignet sich nicht für Daten, bei denen die korrekte Wiederherstellung unabdingbar ist. *Beispiele:* Quellcode, ausführbare Programme, Textdokumente

Aufgabe 4

- *Verlustfrei:* Portable Network Graphics (PNG)
- *Verlustbehaftet:* Joint Photographic Experts Group (JPEG)

Aufgabe 5

RLE, Steuerzeichen: '!', Zählziffer: 4–9

- (a) unkomprimiert: AAAAAAABB!CCCCCCCCCCCCCCC
komprimiert: A!7BB!0C!9C!7

(b) Kompressionsrate: $\frac{|\text{unkomprimierte Daten}|}{|\text{komprimierte Daten}|} = \frac{26}{13} = 2$

Aufgabe 6

RLE, Steuerzeichen '?', eine Zählziffer

- (a) komprimiert: X?04Y?8?0ZZZ
unkomprimiert: X?4YYYYYYYY?ZZZ

(b) Kompressionsrate: $\frac{|\text{unkomprimierte Daten}|}{|\text{komprimierte Daten}|} = \frac{26}{13} = 2$

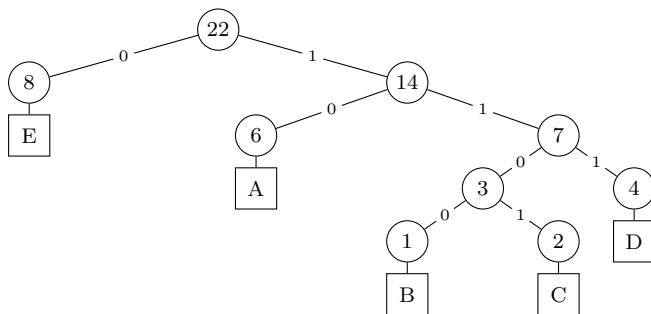
Aufgabe 7

- (a) $C = \{01, 001, 100, 101, 110\}$ präfix-frei
- (b) $C = \{01, 10, 11, 111\}$ nicht präfix-frei
- (c) $C = \{000, 001, 010, 011, 100, 101, 110, 111\}$ präfix-frei

Aufgabe 8

BCCDDDDAAAAAAEEEEEEEE \Rightarrow

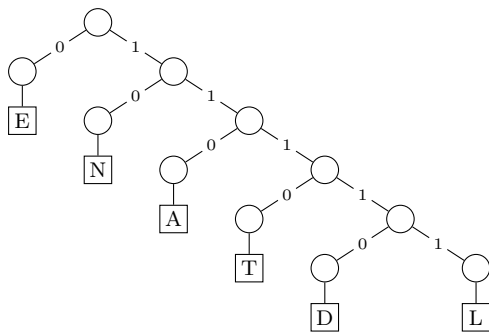
Zeichen	B	C	D	A	E
Häufigkeit	1	2	4	6	8



Zeichen	A	B	C	D	E
Code	10	1100	1101	111	0

$$\text{Kompressionsrate} = \frac{21 \cdot 3}{1 \cdot 4 + 2 \cdot 4 + 4 \cdot 3 + 6 \cdot 2 + 8 \cdot 1} = \frac{63}{44} \approx 1.43$$

Aufgabe 9



(a)

1110	110	10	10	0	10	10	110	11110	0	11111	10
T	A	N	N	E	N	N	A	D	E	L	N

(b)

Zeichen	N	E	A	T	L	D
Häufigkeit	5	2	2	1	1	1

Für TANNENADELN ist der Baum nicht optimal, da N häufiger als E vorkommt aber keinen kürzeren Code als E hat.

Aufgabe 10

Unkomprimiert: AAAAAAAAAABC

Basistabelle: A=0, B=1, C=2

Substr	Code	Tab+
A	0	AA=3
AA	3	AAA=4
AAA	4	AAAA=5
AAAA	5	AAAAB=6
B	1	BC=7
C	2	--

Komprimiert: 0, 3, 4, 5, 1, 2

Anzahl Bits unkomprimiert: $12 \cdot 2 = 24$

Anzahl Bits komprimiert: $6 \cdot 3 = 18$

Kompressionsrate: $24/18 = 4/3 = 1.33$

Aufgabe 11

Basistabelle: A=0, O=1, R=2

Code	Zeichen	Tabelle+
2	R	3=RO
1	O	4=OA
0	A	5=AR
2	R	6=RR
6	RR	7=RRR
7	RRR	--

Ein im letzten Schritt erzeugter Tabelleneintrag kann nur dann im folgenden Schritt verwendet werden, wenn das (unbekannte) anzuhängende Zeichen (rot) mit dem ersten Zeichen der letzten eingefügten Zeichenfolge (blau) übereinstimmt.

Somit wurde ROARRRRRR übertragen.

Bei einer Zeichenkette aus 10 Zeichen mit 3 verschiedenen Symbolen werden ohne Kompression $10 \cdot 2 \text{ Bit} = 20 \text{ Bit}$ benötigt.

In der komprimierten Form wurden 6 Zahlen mit jeweils 3 Bit übertragen und somit 18 Bit benötigt.

Also beträgt die Kompressionsrate $20/18 = 10/9 = 1.11$