

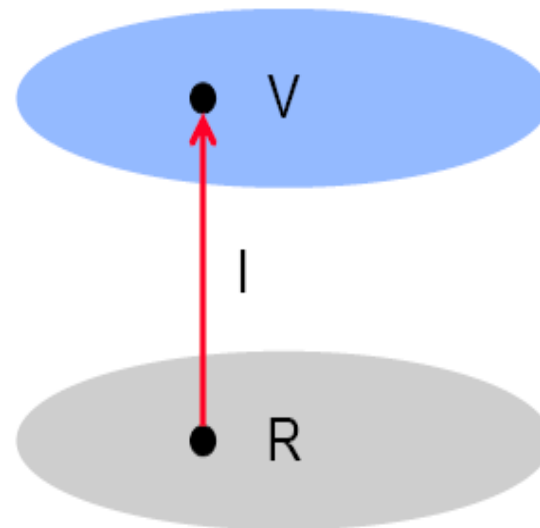
Übungsblätter

- 13 Übungsblätter
- Ab dem 3. praktische Arbeit am PC
- 5 Hausaufgaben
- 1-4 Punkte pro Hausaufgabe
- 12 von 20 Punkten zur Klausurzulassung

Darstellung von Zeichen und Zahlen

Information und Repräsentation

- V = Menge von *Werten* (Interpretationen, Bedeutungen)
- R = Menge von *Repräsentationen* (Darstellungswerten)
- Abbildung
 $I : R \rightarrow V$ *Interpretation*
- Umkehrung zur Interpretation: *Repräsentationsbeziehung* $I^{-1} : V \rightarrow R$



Klassische Beispiele:

V = Zahlwerte, R = Binärzahlen

V = Abbildungen, R = Programme

Hier betrachtete Beispiele:

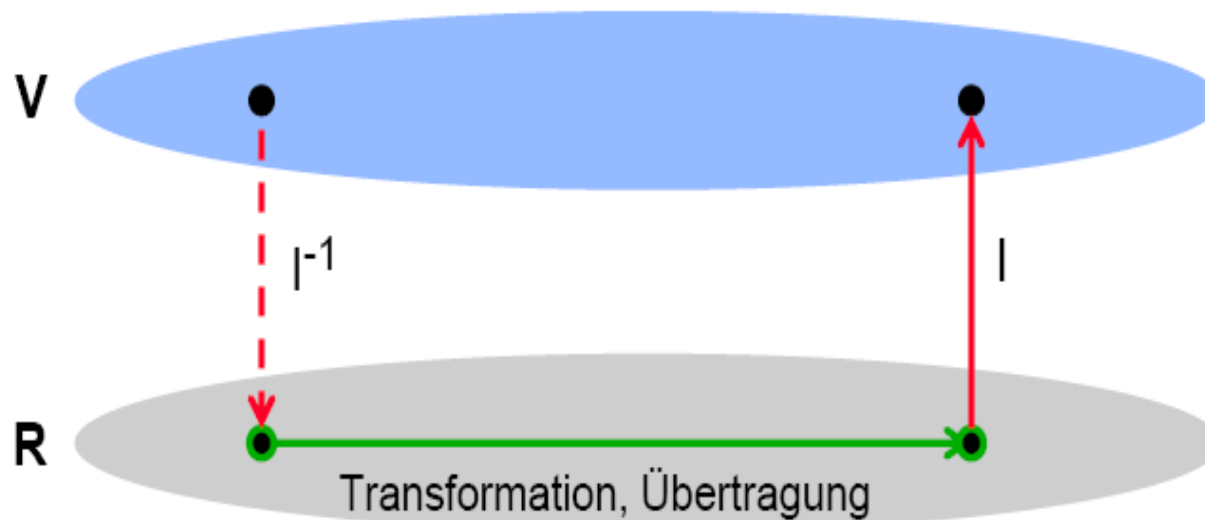
V = textuelle Aussagen (z.B.)

R = Bilder, R = Klänge, ...

(nach Broy: Informatik Teil I)

Informationsverarbeitung

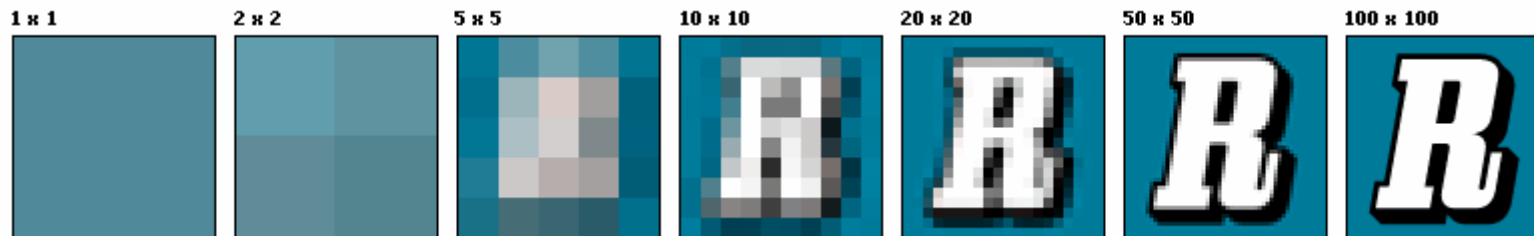
- Information ist ein abstrakter Begriff.
- Computer verarbeiten immer Repräsentationen.
- Informationsverarbeitung ist Repräsentationsverarbeitung.
- Medien sind spezielle Repräsentationen von Information.



Speicherplatz

Aufgabe a)

- $1024 \text{ px} \times 768 = 786432 \text{ px}$
- $786432 \text{ px} * 24 \text{ bit/px} = 18874368 \text{ bit}$
- $18874368 \text{ bit} / 8 = 2359296 \text{ byte}$
- $2359296 / 1024 = 2304 \text{ KByte}$
- $2304 / 1024 = 2,25 \text{ MB}$
- als JPEG ca. 50 KB



Aufgabe b)

- $320 \text{ px} \times 240 = 76800 \text{ px}$
- $76800 \text{ px} * 24 \text{ bit/px} = 1843200 \text{ bit (1 Bild)}$
- $1843200 \text{ bit} * 30 \text{ fps} = 55296000 \text{ bps}$
- $55296000 / 1024 = 54000 \text{ Kbps}$
- $54000 / 1024 \approx 52,73 \text{ Mbps} = 6,6 \text{ MBps}$
- normales Fernsehen 768x576 Pixel

Aufgabe c)

- Aus b) 1 Bild = 1843200 bit
- $15 \text{ fps} * 3 * 60\text{s} = 2700 \text{ Frames}$
- Bildmaterial = $1843200 * 2700 = 593 \text{ MB}$
- Audio: $180 \text{ s} * 22000 \text{ 1/s} * 16 \text{ b} =$
- 7,55 MB
- Zusammen: $593 + 7,55 = \text{ca. } 600 \text{ MB}$
- für 3 Minuten! Ohne Kompression nicht möglich

Laufängencodierung

- Text kürzen, Repräsentation erhalten

AAAAABBBBB

wird mit Hilfe einer Konvention zu

#A5#B5

Aufgabe 2

- 1110 0000 00FF FFFF FFFF 5656 6556 ABBA
CBBB BBBB B773 3333

- wird zu


#13 #07 #F9 F 56566556ABBAC #B8 77 #35

- 60 Zeichen zu 31 Zeichen zusammengefasst

Lauf längencodierung

- Konvention muss genau sein!
- Was passiert beim # Zeichen
- Was passiert bei 20 mal dem selben Zeichen?
- Qualität der Lauf längencodierung schlecht bei 1212121212.
- Abhilfe schafft die Wörterbuchkompression (LZW).

Kompressionsverfahren: Übersicht

- Klassifikationen:
 - Universell vs. speziell (für bestimmte Informationstypen)
 - Verlustfrei vs. verlustbehaftet
 - In diesem Kapitel: nur universelle & verlustfreie Verfahren
- Im folgenden vorgestellte Verfahren:
 - Statistische Verfahren:
 - » Huffman-Codierung 
 - » Arithmetische Codierung
 - Zeichenorientierte Verfahren:
 - » Lauflängencodierung (RLE Run Length Encoding)
 - » LZW-Codierung

Huffman Codierung

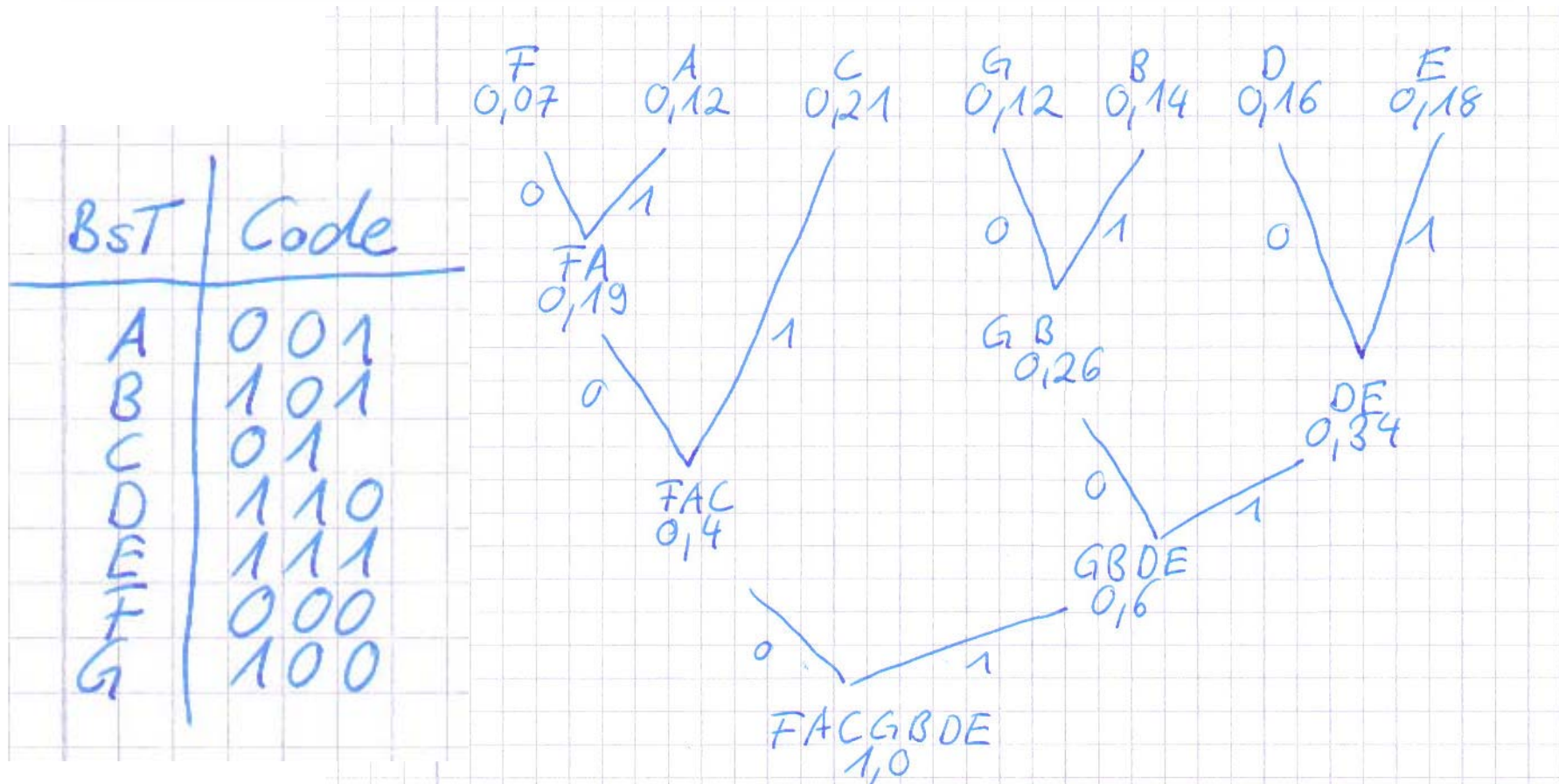
Aufgabe 3

Zeichen	A	B	C	D	E	F	G
Häufigkeit	0,12	0,14	0,21	0,16	0,18	0,07	0,12

- Geben Sie die zugehörige Huffman-Codierung an.
- Berechnen Sie die Redundanz des Codes.
- Codieren Sie die folgenden Zeichenreihe: DCBDAGFD
- Berechnen Sie die Entropie der Nachrichtenquelle.
- Ist der Code optimal?

Code-Baum und Tabelle

Zeichen	A	B	C	D	E	F	G
Häufigkeit	0,12	0,14	0,21	0,16	0,18	0,07	0,12



Aufgabe 3b)

Zeichen	A	B	C	D	E	F	G
Häufigkeit	0,12	0,14	0,21	0,16	0,18	0,07	0,12

$$\text{Redundanz} = L - H$$

durchschnittliche Wortlänge L

$$L = \sum_{a \in A} p_a |c(a)|$$

Entropie

$$H = \sum_{a \in A} p_a \log_2 \left(\frac{1}{p_a} \right)$$

BST	Code
A	001
B	101
C	01
D	110
E	111
F	000
G	100

$$\begin{aligned} L &= 0,12 \cdot 3 \text{ (A)} \\ &+ 0,18 \cdot 3 \text{ (B)} \\ &+ 0,21 \cdot 2 \text{ (C)} \\ &+ 0,16 \cdot 3 \text{ (D)} \\ &+ 0,18 \cdot 3 \text{ (E)} \\ &+ 0,07 \cdot 3 \text{ (F)} \\ &+ 0,16 \cdot 3 \text{ (G)} \\ &= 0,79 \cdot 3 + 0,21 \cdot 2 \\ &= 2,79 \end{aligned}$$

Aufgabe 3b)

- Entropie ganz genau in „Diskrete Strukturen“

Entropie (2)

- Durchschnittlicher Entscheidungsgehalt je Zeichen: Entropie H

$$H = \sum_{a \in A} p_a \operatorname{ld} \left(\frac{1}{p_a} \right)$$

mit $x_a = \operatorname{ld} (1/p_a)$: $H = \sum_{a \in A} p_a x_a$

Quelle 1	Zeichen a	A	B	C	D	$H = 0$
	Häufigk. p_a	1	0	0	0	
	x_a	0	-	-	-	
Quelle 2	Zeichen a	A	B	C	D	$H = 2$
	Häufigk. p_a	0.25	0.25	0.25	0.25	
	x_a	2	2	2	2	
Quelle 3	Zeichen a	A	B	C	D	$H = 1.75$
	Häufigk. p_a	0.5	0.25	0.125	0.125	
	x_a	1	2	3	3	

Entropie ist Maß für „Unordnung“, „Zufälligkeit“

Aufgabe 3b)

Zeichen	A	B	C	D	E	F	G
Häufigkeit	0,12	0,14	0,21	0,16	0,18	0,07	0,12

Entropie

$$H = \sum_{a \in A} p_a \lg \left(\frac{1}{p_a} \right)$$

$$\begin{aligned} H &= 0,12 \cdot 3,059 \text{ (A)} \approx 0,36 \\ &= +0,14 \cdot 2,837 \text{ (B)} + 0,397 \\ &\quad + 0,21 \cdot 2,252 \text{ (C)} + 0,473 \\ &\quad + 0,16 \cdot 2,644 \text{ (D)} + 0,423 \\ &\quad + 0,18 \cdot 2,474 \text{ (E)} + 0,445 \\ &\quad + 0,07 \cdot 3,837 \text{ (F)} + 0,269 \\ &\quad + 0,12 \cdot 3,059 \text{ (G)} + 0,36 \\ &\approx 2,74 \end{aligned}$$

$$R = L - H$$

$$R = 2,79 - 2,74 = \underline{\underline{0,05}}$$

Aufgabe 3c)

Bst	Code
A	001
B	101
C	01
D	110
E	111
F	000
G	100

- **DCBDAGFD**
- wird zu
- 110
- 01
- 101
- 110
- 001
- 100
- 000
- 110

Aufgabe 3e)

- Ist der Code optimal?

$$R = L - H$$
$$R = 2,79 - 2,74 = \underline{\underline{0,05}}$$

- Wann wird der Code optimal?
- Wenn die Auftrittswahrscheinlichkeiten umgekehrte Potenzen von 2 sind. Also $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{16}$ usw.

Aufgabe 4 (optional): Huffman-Codierung

Gegeben sei die folgende Nachricht:

RZROZPZZ

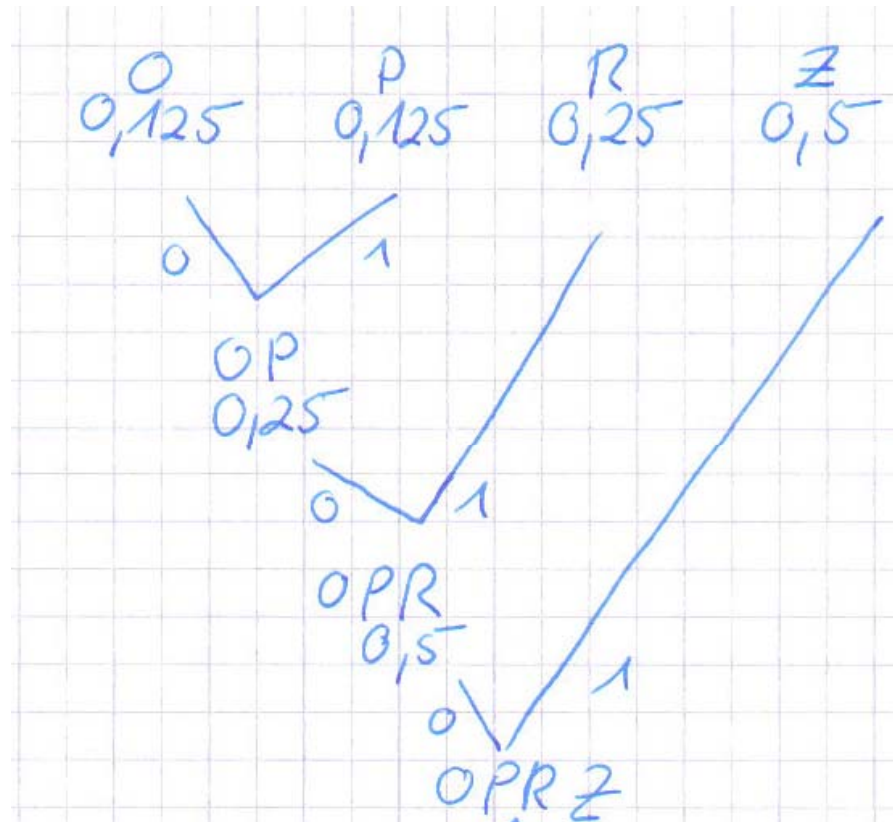
- a) Geben Sie die zugehörige Huffman-Codierung an.
- b) Ermitteln Sie durch Rechnung, ob es sich dabei um einen optimalen Code handelt und begründen Sie das Ergebnis. Welche Regel trifft hier zu?

Aufgabe 4a)

RZROZPZZ

0	1	$\frac{1}{8}$
P	1	$\frac{1}{8}$
R	2	$\frac{1}{4}$
Z	4	$\frac{1}{2}$

BsT	Code
0	000
P	001
R	01
Z	1



Aufgabe 4 (optional): Huffman-Codierung

Gegeben sei die folgende Nachricht:

RZROZPZZ

- a) Geben Sie die zugehörige Huffman-Codierung an.
- b) Ermitteln Sie durch Rechnung, ob es sich dabei um einen optimalen Code handelt und begründen Sie das Ergebnis. Welche Regel trifft hier zu?

Fano-Bedingung!

Huffman Codierung

- Huffman Codierung
 - Vorteil: einfach, schnell
 - Nachteil: nicht immer optimal
- Arithmetische Codierung
 - Vorteil: immer optimal, theoretisch die Beste
 - Nachteil: Patent der Firmen IBM, AT& und Mitsubishi

Aufgabe 5: Arithmetische Codierung

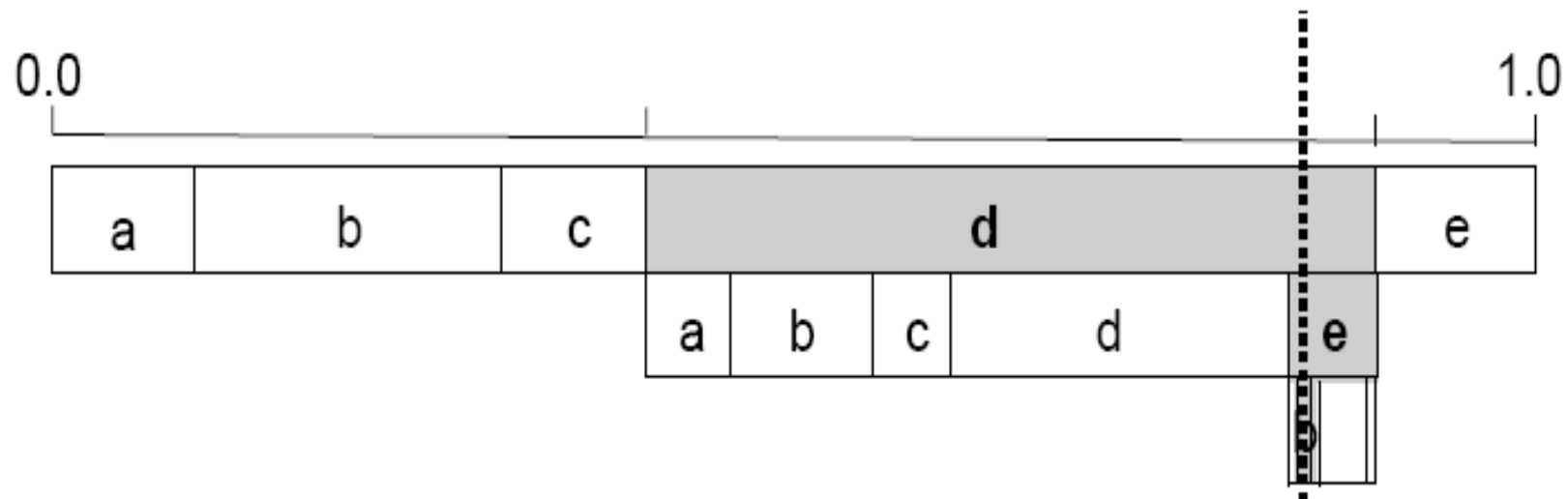
Es sei folgende Nachricht gegeben:

DBBDADBDC

- a) Veranschaulichen Sie die arithmetische Codierung, indem Sie für die ersten 3 Zeichen der Nachricht die Aufteilung in Intervalle graphisch darstellen.
- b) Codieren Sie die gesamte Nachricht mittels des Algorithmus der arithmetischen Codierung.

Arithmetische Codierung (1)

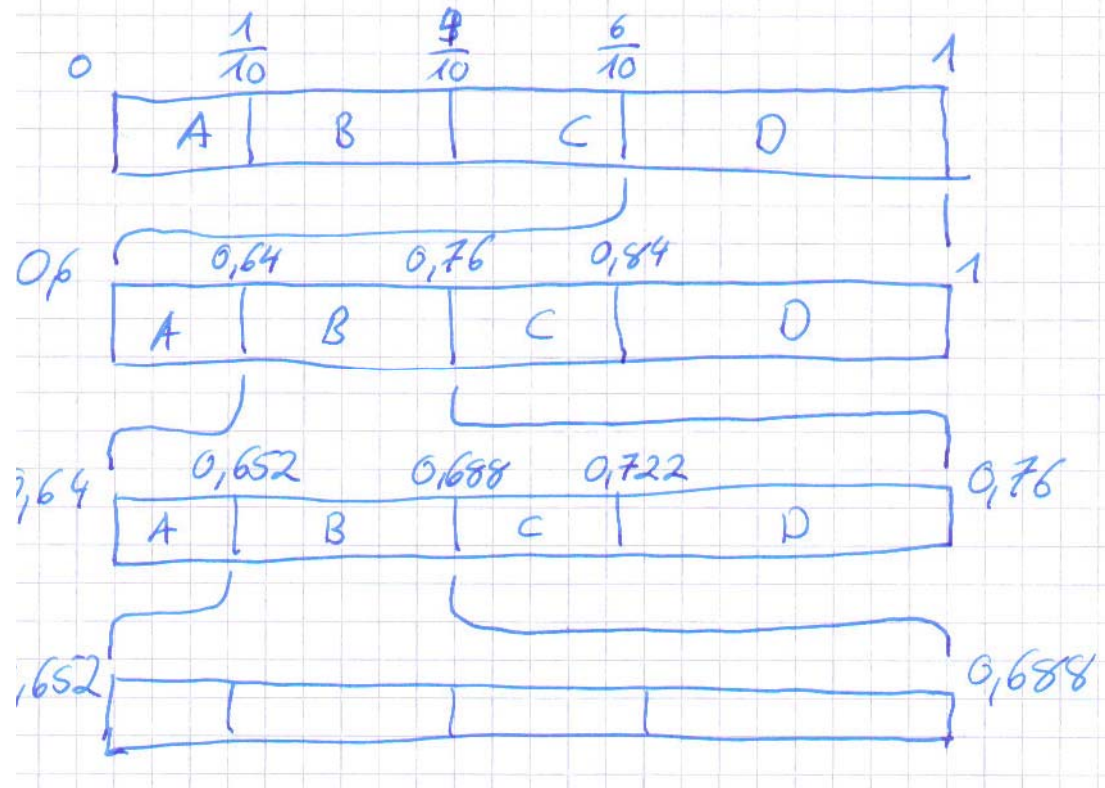
- Gegeben: Zeichenvorrat und Häufigkeitsverteilung
- Ziel: Bessere Eignung für Häufigkeiten, die keine Kehrwerte von Zweierpotenzen sind
- Patentiertes Verfahren; nur mit Lizenz verwendbar
- Grundidee:
 - Code = Gleitkommazahl berechnet aus den Zeichenhäufigkeiten
 - Jedes Eingabezeichen bestimmt ein Teilintervall



Aufgabe 5a)

DBBDADB DCC

A	1	$\frac{1}{10}$
B	3	$\frac{3}{10}$
C	2	$\frac{2}{10}$
D	4	$\frac{4}{10}$



Aufgabe 5b)

real L = 0.0; real R = 1.0;

Solange Zeichen vorhanden **wiederhole**

Lies Zeichen und bestimme Zeichenindex i;

real B = (R-L);

$R = L + B * R_i;$

$L = L + B * L_i;$

Ende Wiederholung;

Code des Textes ist Zahl im Intervall [L, R]

Aufgabe 5b)

Zeichen	A	B	C	D
Häufigk. (p_i)	0,1	0,3	0,2	0,4
linker Rand	0	0,1	0,4	0,6
rechter Rand	0,1	0,4	0,6	1

real L = 0.0; real R = 1.0;

Solange Zeichen vorhanden **wiederhole**

Lies Zeichen und bestimme Zeichenindex i;

real B = (R-L);

R = L + B*R_i;

L = L + B*L_i;

Ende Wiederholung;

Code des Textes ist Zahl im Intervall [L, R]

	B	L	R
		0	1
D	1,0000000000	0,6000000000	1,0000000000
B	0,4000000000	0,6400000000	0,7600000000
B	0,1200000000	0,6520000000	0,6880000000
D	0,0360000000	0,6736000000	0,6880000000
A	0,0144000000	0,6736000000	0,6750400000
D	0,0014400000	0,6744640000	0,6750400000
B	0,0005760000	0,6745216000	0,6746944000
D	0,0001728000	0,6746252800	0,6746944000
C	0,0000691200	0,6746529280	0,6746667520
C	0,0000138240	0,6746584576	0,6746612224

Codiert: 0,67466