Informatik IV
im Sommersemester 2002

IBDS Universität Karlsruhe (TH)
Prof. Dr. Alfred Schmitt
Dipl.-Inform. B. Klimmek
Gebäude 50.34, Am Fasanengarten 5

Musterlösung zur Klausur vom 19. Oktober 2002

Aufgabe 1

Geben Sie für $n \geq 5$ die Wahrscheinlichkeit $p_n(n, 5)$ an, daß bei gleichverteiltem Hashing von n Schlüsseln in n Zellen in der ersten Zelle genau 5 Schlüssel zu liegen kommen.

$$p_n(n,5) = \binom{n}{5} p^5 \, (1-p)^{n-5} \qquad \text{nach Bernoulli, mit } p = \frac{1}{n}$$

$$= \binom{n}{5} \frac{1}{n^5} \cdot \frac{(n-1)^{n-5}}{n^{n-5}} \quad = \quad \binom{n}{5} \frac{(n-1)^{n-5}}{n^n}$$

(Anmerkung: Statt “in der ersten Zelle” hätte es auch heißen können “in einer bestimmten Zelle”, da wir es mit gleichverteiltem Hashing zu tun haben; die Nummer der Zelle geht nicht in die Rechnung ein.)

5 Punkte

Aufgabe 2

a) Zeigen Sie, daß der Einheitswürfel $W = [0\,,1]^3$ eine rationale Menge im $\mathbb{R}^3$ ist, indem Sie ihn in die Standardform für rationale Mengen überführen.

Sei ein Vektor im $\mathbb{R}^3$ bestehend aus den Komponenten $X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$. Dann ist

$$W \;=\; \{\; X \in \mathbb{R}^3 \;\mid\; B_1(X) \geq 0 \wedge B_2(X) \geq 0 \wedge B_3(X) \geq 0 \wedge B_4(X) \geq 0 \wedge B_5(X) \geq 0 \wedge B_6(X) \geq 0 \;\}$$

mit

$$B_1(X) = x \quad \text{und} \quad B_2(X) = 1-x \qquad \text{wegen} \quad 0 \leq x \leq 1$$

$$B_3(X) = y \quad \text{und} \quad B_4(X) = 1-y \qquad \text{wegen} \quad 0 \leq y \leq 1$$

$$B_5(X) = z \quad \text{und} \quad B_6(X) = 1-z \qquad \text{wegen} \quad 0 \leq z \leq 1$$

(4 Punkte)

b) Definieren Sie im $\mathbb{R}^3$ eine rationale Menge mit unendlich vielen Elementen, die dünn ist. Begründen Sie alle drei Eigenschaften kurz.

$$M \quad = \quad \{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \;\mid\; z = 0 \,\}$$

- $|M| = \infty$, denn M ist eine (Hyper-)Ebene.
- M ist rational, denn $B_1(X) = z \geq 0$ und $B_2(X) = -z \geq 0$.
- Keine ε-Kugel paßt in M, denn die Kugel hat (für $\varepsilon > 0$) auch eine z-Ausdehnung; also ist M dünn.

(Anmerkung: Jedes ein- oder zweidimensionale Artefakt im $\mathbb{R}^3$ ist hier geeignet, beispielsweise die Gerade $M = \{\, (x\; y\; z)^\top \in \mathbb{R}^3 \;\mid\; x = y = 0 \,\}$.)

(3 Punkte)

7 Punkte

Aufgabe 3

Skizzieren Sie ein neuronales Netz, bestehend aus nur einem Neuron, das die logische Funktion

$$f(A, B, C) \quad = \quad (A \vee B) \wedge (B \vee C) \wedge (C \vee A)$$

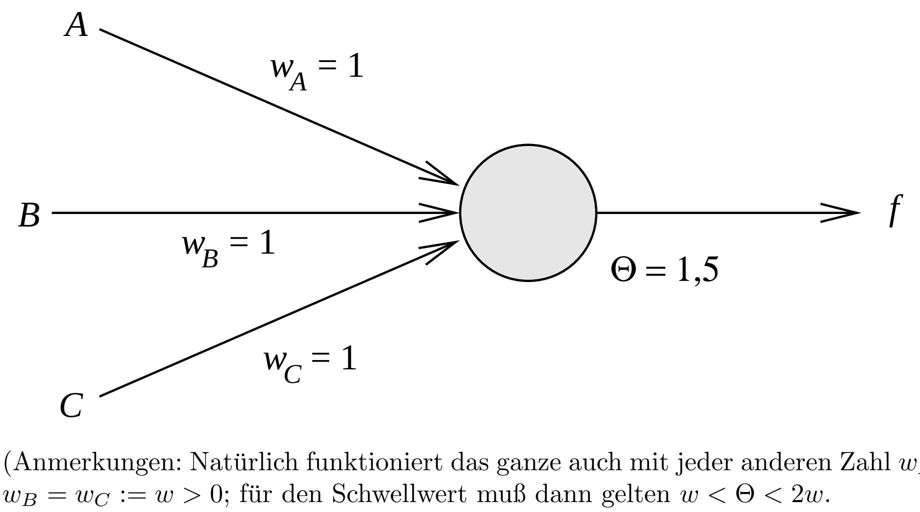
realisiert. Geben Sie die Gewichte w_A, w_B, w_C der Eingaben sowie den Schwellwert Θ des Neurons an, wenn sowohl für die Eingaben als auch für die Ausgabe gilt

$$\texttt{TRUE} \equiv 1 \qquad \text{und} \qquad \texttt{FALSE} \equiv 0$$

(Das Neuron sei dabei folgendermaßen beschaffen: ist die gewichtete Summe der Eingaben $w_A A + w_B B + w_C C$ größer als der Schwellwert Θ, so sei die Ausgabe Eins; ist $w_A A + w_B B + w_C C < \Theta$, so sei die Ausgabe Null; andernfalls ist die Ausgabe als undefiniert anzusehen!)

A	B	C	f	$A+B+C$
0	0	0	0	0
0	0	1	0	1
0	1	0	0	1
0	1	1	1	2
1	0	0	0	1
1	0	1	1	2
1	1	0	1	2
1	1	1	1	3

Wie aus der Tabelle zu ersehen ist, ist f genau dann TRUE (also Eins), wenn $A + B + C \geq 2$. Setze also alle Gewichte $w_A = w_B = w_C := 1$ und die Schwelle auf $\Theta := 1\frac{1}{2}$.



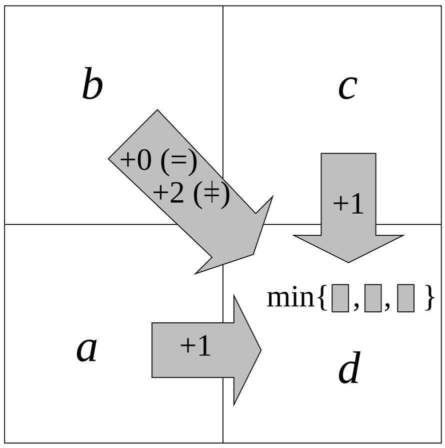
(Anmerkungen: Natürlich funktioniert das ganze auch mit jeder anderen Zahl $w_A = w_B = w_C := w > 0$; für den Schwellwert muß dann gelten $w < \Theta < 2w$.
Anstatt der Wertetabelle hätte man auch mit De Morgan die Formel umformen können in $f = (A \wedge B) \vee (B \wedge C) \vee (C \wedge A)$; hier kann quasi direkt die Schwelle abgelesen werden.)

8 Punkte

Aufgabe 4

Betrachten Sie den Editierabstand von Zeichenketten, wenn als einzige Elementaroperation das Zeichen-Löschen (*delete*) erlaubt ist. Um also den Abstand der beiden Zeichenketten zu bestimmen, sollen möglichst wenige Zeichen in den Zeichenketten gelöscht werden, um Gleichheit der beiden übrigbleibenden Zeichenketten zu erreichen.

a) Mit Hilfe einer $m \times n$-Matrix soll der Editierabstand zweier Zeichenketten v und w mit m bzw. n Buchstaben bestimmt werden. Geben Sie ein Schema an, nach dem ein Matrixelement (i, j) aus benachbarten Matrixelementen ermittelt werden kann.



Mit $a = M(i, j-1)$, $b = M(i-1, j-1)$, $c = M(i-1, j)$ und

$$b' = \begin{cases} b & \Leftarrow \text{ gleiches Zeichen übernehmen } \Leftarrow v[i] = w[j] \\ b+2 & \Leftarrow \text{ in } v \text{ und } w \text{ löschen } \Leftarrow v[i] \neq w[j] \end{cases}$$

ist

$$d = M(i, j) = \min\{ \underbrace{a+1}_{\text{1 Zeichen in } w \text{ löschen}}, b', \underbrace{c+1}_{\text{1 Zeichen in } v \text{ löschen}} \}$$

(4 Punkte)

b) Bestimmen Sie den Editierabstand zwischen `ABAABB` und `AABBAB`:

		A	A	B	B	A	B
	0	1	2	3	4	5	6
A	1	**0**	**1**	2	3	**4**	5
B	2	1	2	**1**	**2**	3	**4**
A	3	**2**	**1**	2	3	**2**	3
A	4	**3**	**2**	3	4	**3**	4
B	5	4	3	**2**	**3**	4	**3**
B	6	5	4	**3**	**2**	3	**4**

Das Endergebnis lautet also: 4 Editieroperationen.

(4 Punkte)

8 Punkte

Aufgabe 5

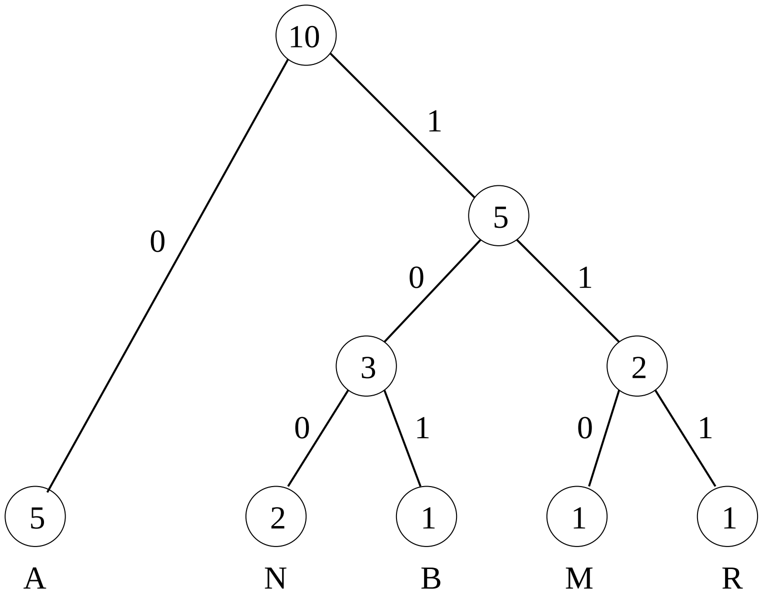
Die Zeichenkette `BANANARAMA` soll mithilfe eines (speziell für die darin vorkommenden Zeichen `A`,`B`,`M`,`N`,`R` zu bestimmenden) binären Huffman-Kodes platzsparend kodiert werden.

a) Bestimmen Sie für jedes dieser Zeichen eine solche Kodierung.

Häufigkeitstabelle:

A	B	M	N	R	Σ
5	1	1	2	1	10

Daraus läßt sich z. B. der folgende Huffman-Kodierungsbaum – bottom-up – erhalten, indem sukzessiv die am seltensten vorkommenden Zeichen bzw. Unterbäume (gegeneinander mit "0" und "1") kodiert werden:



Demnach wäre ein Huffman-Kode also

A	B	M	N	R
0	101	110	100	111

(6 Punkte)

b) Wie viele Bits enthält die Kodierung der Zeichenkette `BANANARAMA` demnach?

Die Kodierung von `BANANARAMA` wäre gemäß (a) die Bitfolge `10101000100011101100`, welche genau 20 Bits enthält.

(Anmerkung: Dieses Ergebnis ist eindeutig bestimmt, insbesondere unabhängig vom verwendeten Huffmankode. So gibt es auch Kodes, die den Buchstaben `N` mit nur 2 Bits kodieren, dafür aber zwei der Buchstaben `B`,`M`,`R` mit 4 Bits.)

(2 Punkte)

c) Bestimmen Sie die absolute Redundanz der Kodierung, indem Sie (nach Shannon) die Realinformation H der ursprünglichen Zeichenkette mit der Nominalinformation/Kodelänge $\widetilde{H}$ ihrer binären Kodierung verrechnen.

Eine Tabelle für die numerischen Werte von $-p \cdot \log p$ finden Sie im Anhang.

$$
\begin{aligned}
\widetilde{H} &= 20 \text{ bit} \qquad (\text{ siehe (b) }) \\
H &= 10 \cdot H\left(\frac{5}{10}, \frac{1}{10}, \frac{1}{10}, \frac{2}{10}, \frac{1}{10}\right) \qquad \leftarrow \text{ in den Klammern die } p_i \\
&= 10 \cdot \sum_i -p_i \log p_i \\
&= 10 \cdot (0{,}500 + 3 \cdot 0{,}332 + 0{,}464) \qquad \text{– siehe Tabelle –} \\
&= 10 \cdot 1{,}960 \qquad \text{leicht schriftlich zu berechnen ...} \\
&= 19{,}6 \text{ bit}
\end{aligned}
$$

Die (absolute) Redundanz berechnet sich als Differenz aus Nominal- und Realinformation zu

$$
\begin{aligned}
R &= \widetilde{H} - H \\
&= 20 \text{ bit} - 19{,}6 \text{ bit} \\
&\approx \frac{1}{2} \text{ bit}
\end{aligned}
$$

(7 Punkte)

15 Punkte

Aufgabe 6

Kodieren Sie die Zeichenkette `ABACDBEACDFGDBEACDFG` nach dem Lempel-Ziv-Welch-Verfahren (LZW).

a) Lassen Sie dabei die neuen Wörterbuch-Einträge bei Kodenummer 256 anfangen.

Eingabe	**Ausgabe**	**Wörterbuch**
A	A	AB 256
B	B	BA 257
A	A	AC 258
C	C	CD 259
D	D	DB 260
B	B	BE 261
E	E	EA 262
A	258	ACD 263
C		
D	D	DF 264
F	F	FG 265
G	G	GD 266
D	260	DBE 267
B		
E	262	EAC 268
A		
C	259	CDF 269
D		
F	265	
G		

(5 Punkte)

b) Wieviel Byte werden eingespart, wenn die nichtkomprimierte Zeichenkette mit 8 bit/Zeichen und die komprimierte mit 9 bit/Ausgabezeichen dargestellt wird?

$$\begin{aligned} \text{unkodiert:}\quad 8 \cdot 20 &= 160 \text{ bit} \\ \text{kodiert:}\quad 9 \cdot 15 &= 135 \text{ bit} \\ \hline & \ 25 \text{ bit gespart} \end{aligned}$$

$$\text{Also } \left\lfloor \frac{25}{8} \right\rfloor = 3 \text{ Byte gespart}$$

(2 Punkte)

7 Punkte

Aufgabe 7

Kreuzen Sie für folgende Aussagen an, ob diese wahr oder falsch sind.
Hinweis: Für jede richtige Antwort gibt es einen Punkt, für jede falsche Antwort wird ein Punkt abgezogen. Die gesamte Aufgabe wird mit einer nichtnegativen Punktezahl gewertet.

Aussage	*Antwort*
Die Vereinigung zweier rationaler Dünn-Mengen ist stets in einer rationalen Dünn-Menge enthalten.	**wahr**
Zweidimensionale Zellraster sind ungeeignet, um sämtliche Berührungs- und Schnittpunkte von n Strecken in der Ebene zu bestimmen.	**falsch**
Linear verkettete AVL-Bäume sind sehr gut geeignet, um Dateien index-sequentiell zu organisieren und auf Platten zu speichern (ISAM-Dateiorganisation).	**falsch**
Invertierte Dateien erlauben eine schnelle Suche nach beliebig vorgegebenen Suchbegriffen.	**wahr**
Udi Manbers Technik *Glimpse* ist eine der besten Methoden, um eine Suchmaschine für das Internet zu implementieren.	**falsch**
Ein störungsfreier Kanal ist deterministisch, aber eventuell verlustbehaftet.	**falsch**
Die im Beweis zum Shannonschen Fundamentalsatz der Informationstheorie vorkommenden Mengen A der typischen Input-Output-Paare sind stets von endlicher Mächtigkeit.	**wahr**
Es gibt optimale Kodes (variabler Länge), die eindeutig entzifferbar, aber nicht sofort entzifferbar sind.	**wahr**
Eine nicht komprimierte Computeranimation von 1 Minute Laufzeit kann 2 Gigabyte Speicher benötigen, wenn sie im Standardformat von $768 \cdot 576$ Pixel pro Bild aufgezeichnet wird.	**wahr**
Polynome in Stützstellendarstellung können nicht schneller als mit FFT (*Fast Fourier Transform*) multipliziert werden.	**falsch**

10 Punkte

Konzeptpapier

Logarithmentabelle (zu Aufgabe 5)

p	$-p \cdot \log_2 p$
0	0,000
1/10	0,332
2/10	0,464
3/10	0,521
4/10	0,529
5/10	0,500
6/10	0,442
7/10	0,360
8/10	0,258
9/10	0,137
1	0,000