Optimale Codes

Nachricht in einem Quellalphabet $\Sigma_1 = \{a, b, c, A, \sqcup, !, \ldots\}$

Codealphabet = $\{0,1\}$

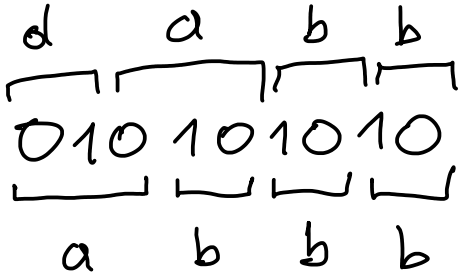
feste Länge 8 Bits

ASCII $\quad a \rightarrow 01100001$

Gesucht ist ein Code $C: \Sigma_1 \rightarrow \{0,1\}^*$

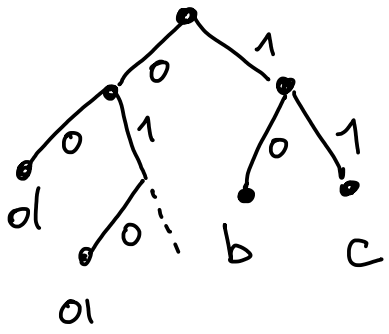
- eindeutig entzifferbar:
  Jede codierte Nachricht $\in \{0,1\}^*$ darf sich auf höchstens eine Art in Codewörter zerlegen lassen.

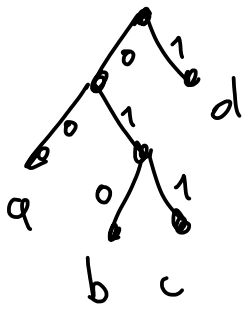
$C_a = 010$
$C_b = 10$
$C_c = 11$
$C_d = 0100$



- stärkere Forderung: präfixfrei.
  Kein Codewort ist Präfix eines anderen Codewortes

Codebaum





$\ell_a = 2 \quad p_a = 20$
$\ell_b = 3 \quad p_b = 17$
$\ell_c = 3 \quad p_c = 14$
$\ell_d = 1 \quad p_d = 77$

$\frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{2} = 1$

Gegeben: Häufigkeiten $p_1, p_2, \dots, p_n$ für die n Zeichen $\in \Sigma$

Gesucht: Ein präfixfreier Code mit Codewortlängen

$$\ell_1, \ell_2, \dots, \ell_n$$

Gesamtlänge $\sum_{i=1}^{n} p_i \ell_i \to \text{MIN}$

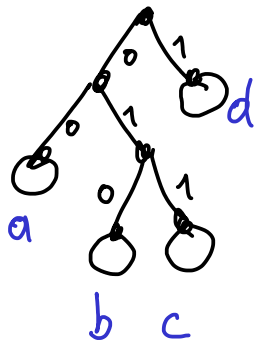
Äquivalent: Ein <u>Codebaum</u> mit n Blättern auf Tiefe $\ell_1, \ell_2, \dots, \ell_n$.

- $p_1, \dots, p_n$ können auch <u>relative Häufigkeiten</u> / Wahrscheinlichkeiten sein.

Algorithmus von Huffman:

① $p_i < p_j \Rightarrow \ell_j \leq \ell_i$

Bew. Ann. $p_i < p_j$ $\ell_j > \ell_i$



$\ell_1 = 1$ $p_a = 20$

$\ell_2 = 2$ $p_b = 17$

$\ell_3 = 3$ $p_c = 14$

$\ell_4 = 3$ $p_d = 77$

$\Sigma' = p_i \ell_i + \dots + p_j \ell_j$

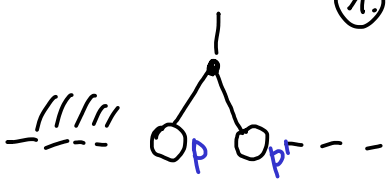
$\Sigma'' = p_i \ell_j + \dots + p_j \ell_i$

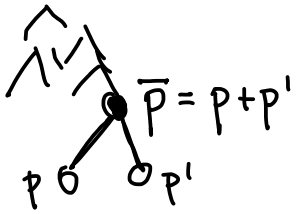
$\Sigma' - \Sigma'' = \dots > 0$

①b. Ein optimaler Codebaum muss <u>voll</u> sein: Jeder innere Knoten hat 2 Kinder.

② In jedem <u>vollen</u> Binärbaum gibt es zwei Geschwisterknoten i, j mit $\ell_i = \ell_j \geq \ell_k$ für alle anderen k (auf der tiefsten Ebene)

①+② p, p' seien die kleinsten Häufigkeiten.

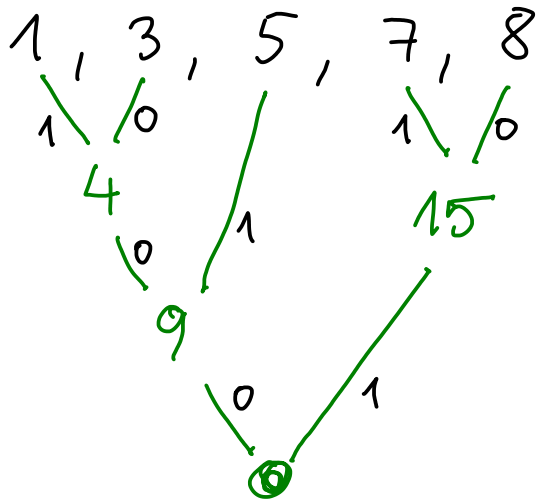




Wir können annehmen, dass p und p' zwei Geschwisterknoten zugeordnet werden.

Ersetze p, p' durch $\bar{p} = p + p'$.

Konstruiere einen opt. Baum für die $n-1$ entstehenden Häufigkeiten.



1: 001
3: 000
5: 01
7: 11
8: 10

Implementierung: Prioritätswarteschlange $O(n \log n)$

Die Kraft-MacMillan'sche Ungleichung

$$(*) \quad \sum_{i=1}^{n} 2^{-\ell_i} \leq 1$$

ist notwendig und hinreichend für

(a) die Existenz eines Binärbaums mit Blättern auf Tiefe $\ell_1, \dots \ell_n$ (vgl. Aufgabe 44)

(b) die Existenz eines präfixfreien Codes mit Wortlängen $\ell_1, \dots, \ell_n$

(c) die Existenz eines ~~präfixfreien~~ eindeutig entzifferbaren Codes mit Wortlängen $\ell_1, \dots, \ell_n$

Satz von McMillan (1956) $(c) \Rightarrow (*)$

$C = \{00, 0101, 1111, 0111, 001100, 010010\}$

$\ell_1 = 2 \quad \ell_2 = 4 \quad \ell_3 = 4 \quad \ell_4 = 4 \quad \ell_5 = 6 \quad \ell_6 = 6$

$\frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^4} + \frac{1}{2^4} + \frac{1}{2^6} + \frac{1}{2^6} \leq 1$

$L = C^*$ $\quad A_k = \left| L \cap \{0,1\}^k \right|$ = Anzahl der Nachrichten der Länge k, die sich aus den Codewörtern bilden lassen

$A_k \leq 2^k$

$A_0 = 1$

$A_k = 0, \quad k < 0$

$$A_k = A_{k-2} + 3 \cdot A_{k-4} + 2 A_{k-6} \quad (k \geq 1)$$

$$A_k = \sum_{i=1}^{n} A_{k-\ell_i}$$

$\left[A_k \sim y_0^k \right]$

Annahme: $1 < \sum_{i=1}^{n} 2^{-\ell_i}$

Finde $y > 2$ mit: $1 < \sum_{i=1}^{n} y^{-\ell_i}$

$(\ell_1 \leq \ell_2 \leq \dots \leq \ell_n)$

Behauptung: $\bar{A}_k \geq c_0 \cdot y^k$ für $k \geq 0$ und für eine Konstante $c_0 > 0$

vollst. Induktion: $\bar{A}_k = \sum_{i=1}^{n} \bar{A}_{k-\ell_i}$ (für $k \geq \ell_n$)

$$\overset{\text{I.V.}}{\geq} \sum_{i=1}^{n} c_0 y^{k-\ell_i} = c_0 \cdot y^k \underbrace{\sum_{i=1}^{n} y^{-\ell_i}}_{>1} > c_0 y^k$$

Induktionsbasis: Wähle $c_0 > 0$ klein genug sodass $c_0 y^k \leq \bar{A}_k$ für $k = 0, 1, \dots, \ell_n - 1$

$\bar{L} = \{\varepsilon, F, FF, \dots, F^{\ell_1 - 1}\} \cdot C^*$ $\quad \bar{A}_k = \left| \bar{L} \cap \{0,1,F\}^k \right| > 0$

Bsp.: F 00 000 11 00 $\in L$, $\in \bar{L}$

$\bar{A}_k \leq 2^k + 2^{k-1} + 2^{k-2} + \dots 2^{k-\ell_1+1} \leq 2^k \cdot 2$

$c_0 y^k \leq \bar{A}_k \leq 2 \cdot 2^k$

$\left(\frac{y}{2}\right)^k \leq \frac{2}{c_0}$ Widerspruch ($\frac{y}{2} > 1$) $\square$