



Warum widerspricht das nicht der Optimalität von Huffman-Codes?

---

Wir kodieren den Text mit

SF: 00 01 00 01 10 11 = 12 Bits

Huff: 0 10 0 10 110 111 = 12 Bits

→ In einem solchen Fall gleichen andere Codewörter die längeren Codewörter aus

↳ In diesem Fall sind die durchschnittlichen Codelängen gleich (2 Bits)

↳  $H_0(T) = 1,918$

→ Oft ist Huffman kürzer als SF

↳ Huffman ist in der Praxis besser zu verwenden

---

Entropie:

$$H_0(T) = \sum_{i=0}^{\sigma-1} \frac{n_i}{n} \lg\left(\frac{n}{n_i}\right)$$
$$H_k(T) = \frac{1}{n} \sum_{w \in \Sigma^k} |T_w| \cdot H_0(T_w)$$

↳ Konkatenation aller Zeichen in  $T$  die auf  $w$  folgen

↳  $H_k$  kann einfach mit Suffix-Trees  
berechnet werden

↳ Hier  $k \leq 3$ , d.h. wir können  $H_k$   
naiv berechnen  $\rightarrow$  Histogramme für  
 $256^3$  Einträge