

Informatik OTG	Codierungstheorie	
	Informationsgehalt, Entropie, mittl. Wortlänge, Redundanz	

Wie viel Information enthält eine diskrete Nachricht, die von einem Sender an einen Empfänger übermittelt wird?

In einem deutschen Text kann man die Auftrittswahrscheinlichkeiten  $p_i$  der einzelnen Buchstaben bestimmen, zum Beispiel  $p_e = 0,1470$  oder  $p_x = 0,0001$ . Das e tritt also viel häufiger als das x auf.

**Informationsgehalt** = Logarithmus zur Basis 2  $\left( \frac{1}{\text{Wahrscheinlichkeit, mit der ein Zeichen auftritt}} \right)$

$$h = \text{ld} \left( \frac{1}{p} \right)$$

Der Informationsgehalt einer Nachricht soll gleich der Summe der Informationsgehalte der einzelnen Zeichen dieser Nachricht sein und außerdem nur von der Wahrscheinlichkeit, mit der das Zeichen gesendet wird, abhängen.

Beispiel: Alphabet {a|b|c}

	p
a	0,50
b	0,25
c	0,25

Berechnen des Informationsgehaltes h:

$$h_a = \text{ld} \frac{1}{0,50} = -\text{ld} 0,50 = 1$$

$$h_b = \text{ld} \frac{1}{0,25} = -\text{ld} 0,25 = 2$$

$$h_c = \text{ld} \frac{1}{0,25} = -\text{ld} 0,25 = 2$$

	$p_i$	$h_i$
a	0,50	1
b	0,25	2
c	0,25	2

ld ist *logarithmus dualis*  
<http://rechneronline.de/logarithmus/>

**Entropie** - mittlerer Informationsgehalt des Ausgangsalphabets

$$H = \sum_i p_i * h_i = \sum_i p_i * \text{ld} \left( \frac{1}{p} \right) = -\sum_i p_i * \text{ld} p_i \quad (\text{Shannonsche Funktion})$$

$$H = 0,5 * 1 + 0,25 * 2 + 0,25 * 2 = 1,5 \text{ Bit/Zeichen}$$

$H \leq L$  (Die Entropie ist maximal, wenn alle Zeichen mit der gleichen Häufigkeit auftreten, dann also  $H = L$ )

**Mittlere Wortlänge** L der gewählten Codierung, (hängt von der Art der gewählten Codierung ab.)

$$L = \sum_i p_i * l_i \quad l_i = \text{Anzahl der Bits des Codes zu einem i-ten Zeichen}$$

	$p_i$	$h_i$	Code	$l_i$
a	0,50	1	1	1
b	0,25	2	01	2
c	0,25	2	00	2

$$L = 0,50 * 1 \text{ Bit} + 0,25 * 2 \text{ Bits} + 0,25 * 2 \text{ Bits} = 1,5 \text{ Bit/Zeichen}$$

**Code-Redundanz** Sie gibt an, wie groß der Anteil der Nachricht ist, der im statistischen Sinne keine Information trägt.

$$R = L - H$$

	$p_i$	$h_i$	$p_i * h_i$	$l_i$	$p_i * l_i$
a	0,7	0,515	0,360	1	0,7
b	0,2	2,322	0,464	2	0,4
c	0,1	3,322	0,332	2	0,2