

Compressão estatística

- Baseada na estimativa (ou cálculo) da frequência de cada símbolo.
 - Símbolos mais frequentes usarão menos bits
 - Símbolos menos frequentes usarão mais bits
- Exemplo:
 - aaaeabbbaaaabcaaadacaaabbbaaaabbcaaaeeaaaba
 - a: 26 vezes Código: 1 d: 1 vez Código: 0110
 - b: 8 vezes Código: 00 e: 2 vezes Código: 0111
 - c: 3 vezes Código: 010

Entropia

$$S_x = -\log_2(P_x)$$

Exemplo:

aaaeabbbaaaabcaaadacaaabbbaaaabbcaaaeeaaaba

- x = Símbolo
- S_x = Entropia de x
- P_x = Probabilidade de x

Total: 40 caracteres

$P_a = 25/40 = 0,625$	$S_a = 0,68 * 25$	= 16,95
$P_b = 8/40 = 0,2$	$S_b = 2,32 * 8$	= 18,58
$P_c = 3/40 = 0,075$	$S_c = 3,74 * 3$	= 11,21
$P_d = 1/40 = 0,025$	$S_d = 5,32 * 1$	= 5,32
$P_e = 2/40 = 0,05$	$S_e = 4,32 * 2$	= 8,54

60,70

Entropia

$$S_x = -\log_2(P_x)$$

Símbolo	Probabilidade	Entropia para cada símbolo	Entropia total
U	12/72		
V	18/72		
W	7/72		
X	15/72		
Y	20/72		

Entropia

$$S_x = -\log_2(P_x)$$

Símbolo	Probabilidade	Entropia para cada símbolo	Entropia total
U	12/72	2,584963	31,01955
V	18/72	2,000000	36,00000
W	7/72	3,362570	23,53799
X	15/72	2,263034	33,94552
Y	20/72	1,847997	36,95994