

# Fiche CODO

ce qu'il faut au min pour compresser l'info sans perte (bits/symboles)

Entropie : Quantité minimale pour ne pas perdre de l'info

$$H = -\sum_i (p_i \log_2(p_i))$$

$p_i$  = proba d'un symbole

$$q(m) = \log_2\left(\frac{1}{p(m)}\right)$$

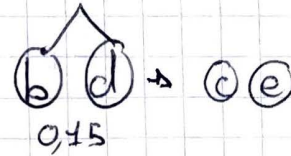
## Huffman

a = 0,2  
b = 0,1  
c = 0,35  
d = 0,05  
e = 0,3

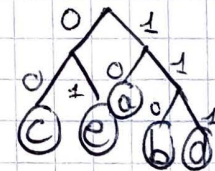
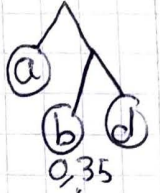
a b c d e

→

a c e



c e



a = 10  
b = 110  
c = 00  
d = 111  
e = 01

$$L_{\text{Huffman}} = \sum p_i \times l_i$$

$$H \leq L_{\text{Huffman}} < H+1$$

## Burrows Wheeler

bactobasic \$  
\$bactobasic  
:

banana \$  
\$banana  
a\$banan  
na\$ban  
ana\$ban  
na\$ba  
anana\$b

\$banana  
a\$banan  
ana\$ban  
na\$ban  
ana\$ban  
na\$ba  
anana\$b

trie

⇒

→ 5

⇒

5annb\$aa

## Move to Front

$\mathcal{L} = \{a, b, n, \$\} = [a, b, n, \$]$

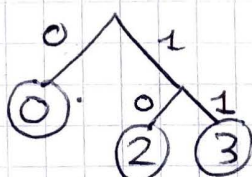
annb\$aa  
↑↑↑↑↑↑

⇒ 0202330

[n, a, b, \$]  
[b, n, a, \$]  
[\$, b, n, a]  
[a, \$, b, n]

bzip2 = Burrows Wheeler → Move To Front → Huffman

0 = 3/7 = 0  
2 = 2/7 = 10  
3 = 2/7 = 11



## Burrows Wheeler decode

	trie	trie	trie
a	a\$	a\$b	a\$b a
n	n a	n a \$	n a \$ b
n	n n	n a n	n a n a
b	b a	b a n	b a n n
\$	\$ b	\$ b a	\$ b a n
a	a n	a n a	a n a \$
a	a n	a n n	a n a n

# LZW:

$L = [a, e, i, p, m, v]$   
 @0...@5

namaPa mananaive

3 bits aPa base

$8 \times 3 + 5 \times 4 = 44$  bits

3 bits

4 bits

Buffer

Input

Dico

Output

n  
a  
n  
i  
e  
p  
a  
n  
a  
n  
a  
n  
a  
i  
v  
e  
\$

@6 = ma  
@7 = an  
@8 = ni  
@9 = ie  
@10 = ep  
@11 = Pa

@12 = ana

@13 = aman

@14 = mai

@15 = iv

@16 = ve

@ m = 4  
@ a = 0  
@ n = 4  
@ i = 2  
@ e = 1  
@ p = 3

@ an = 7

@% @12 a = 12

@ na = 6

@ i = 2

@ v = 5

@ e = 1

On day  
with 4 bits