



## Du binaire compressé

Cette énigme a été créée par **jaudi**

```
11010000110100101001110101000100110010001010111011000011111
01100101000010110010000111110000010011010010110000111001110
10010001111000011101011111110110000000111011110000111010100
01101101001111100000001110111100001111110111001111011111101
11110001000111001011000000111010100011011000001101000011010
100100010110010001010101100111111111111111111010101101011111
111110111111111000110111100001111010001100111111101001010000
11110110010001010000011110110010100011001100001001001111100
010001100101110110001001
```

### Méthode utilisée (indication) :

L'algorithme utilisé permet une compression des données sous format binaire. Inventé en 1952 par David Albert Huffman, il autorise une économie conséquente de 0 et de 1 en adaptant le codage des lettres à leur fréquence (un peu à l'image du Morse). Pour cette énigme, la version avec un dictionnaire statique du codage est utilisée, qui prend en argument la fréquence des lettres en français. On utilisera l'alphabet alpha = « ABCDEFGHIJKLMNOPQRSTUVWXYZ\_ » comprenant 26 lettres et l'espace.

Elle construit l'arbre progressivement en fusionnant les deux nœuds les moins fréquents en un nouveau nœud intermédiaire dont la fréquence vaut la somme de ses deux feuilles.

Une fois que tous les nœuds sont construits, les nœuds sont classés par longueur (correspondant aux nombres de branches les séparant du tronc) puis par ordre alphabétique. Ensuite, le premier symbole dans le classement reçoit un code de longueur n (avec n le nombre de branches le séparant du tronc) et avec que des 0, puis le suivant aura (n-1) fois 0 puis un 1 (puis éventuellement autre chose s'il est plus long que le premier) et ainsi de suite, en veillant à ce qu'aucun des codes n'aient leur début correspondant au code d'une lettre plus courte.

Cette convention, usuellement employée dans les algorithmes, n'est pas toujours celle utilisée sur des sites de cryptographie (notamment Dcode).

### Exemple :

Dans la langue d'alphabet « BACD » (utilisée dans des contrées lointaines), la fréquence des lettres est B : 51 %, A = 15 %, D = 15 % et C = 19 %.

On fusionne A et D pour former nœud\_1 (30 %)

On fusionne C et nœud\_1 pour former nœud\_2 (49 %)

On fusionne B et nœud\_2 pour former nœud\_3 (100 %), qui correspond donc au tronc.

Les lettres dans l'ordre sont donc B (1 bit), C (2 bits), A (3 bits) et D (3 bits).

On en déduit alors le dictionnaire utilisé : B=0, C=10, A=110 et D=111.

Par exemple, BAC sera chiffré 011010 (6 bits, ce qui est un gain énorme par rapport aux  $3*8=24$  bits habituellement nécessaire pour représenter ce code par ASCII).