

Codage de Huffman

Texte à compresser : "le_loup_vole_le_poele".

Caractère	Fréquence
l	5
e	5
_	4
o	3
p	2
u	1
v	1

Attention : cet exemple est volontairement peu efficace en compression afin de voir plusieurs cas de figure en une seule application de l'algorithme.

Codage de Huffman

Caractère	Fréquence
l	5
e	5
-	4
o	3
p	2
u	1
v	1

Codage de Huffman

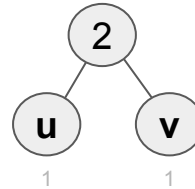
Caractère	Fréquence
l	5
e	5
-	4
o	3
p	2
u	1
v	1

1. Tant que la liste des noeuds n'est pas vide :
 - a. Trier la liste des noeuds par fréquence.
 - b. Sélectionner les deux noeuds les moins fréquents et former un arbre dont la racine prend la valeur de leur somme.
 - c. Insérer l'arbre obtenu dans la liste des noeuds.
2. Pour former les clefs, attribuer la valeur de codage 0 à toutes les branches de gauche, et 1 à toutes les branches de droite.

Codage de Huffman

Caractère	Fréquence
l	5
e	5
-	4
o	3
p	2
u	1
v	1

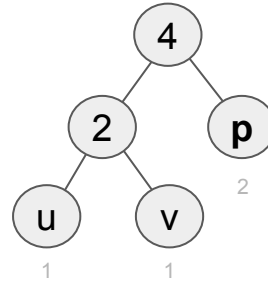
} $1+1=2$



Codage de Huffman

Caractère	Fréquence
l	5
e	5
-	4
o	3
p	2
u, v	2

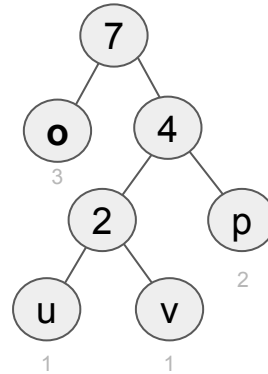
} $2+2=4$



Codage de Huffman

Caractère	Fréquence
l	5
e	5
-	4
u, v, p	4
o	3

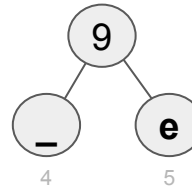
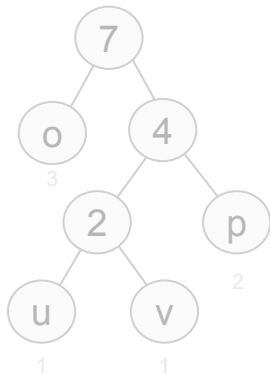
} $3+4=7$



Codage de Huffman

Caractère	Fréquence
u, v, p, o	7
l	5
e	5
-	4

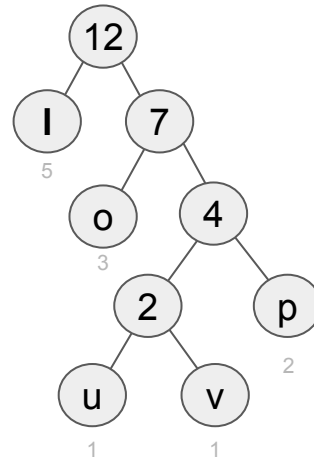
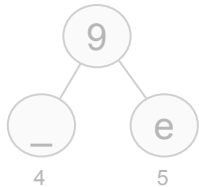
} $4+5=9$



Codage de Huffman

Caractère	Fréquence
e, _	9
u, v, p, o	7
l	5

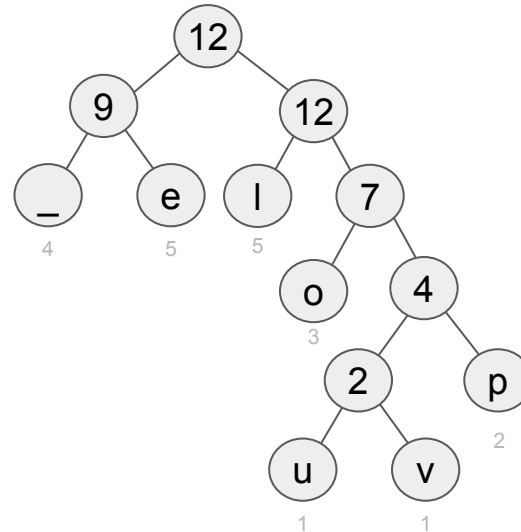
} $5+7=12$



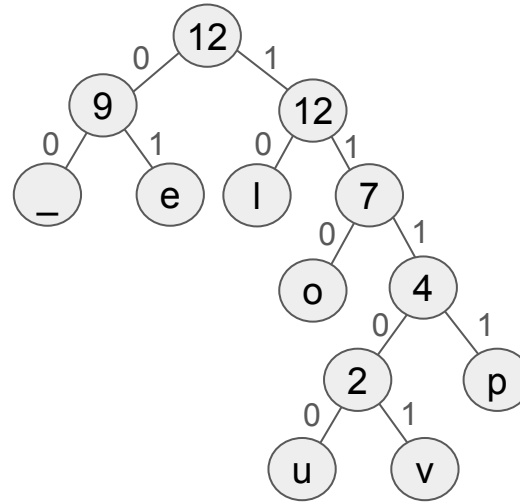
Codage de Huffman

Caractère	Fréquence
u, v, p, o, l	12
e, _	9

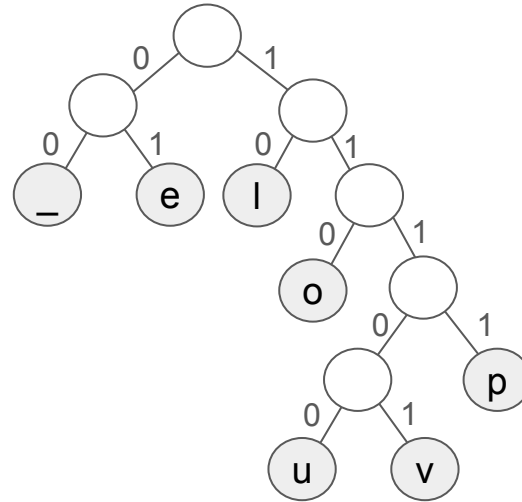
} 9+12=21



Codage de Huffman

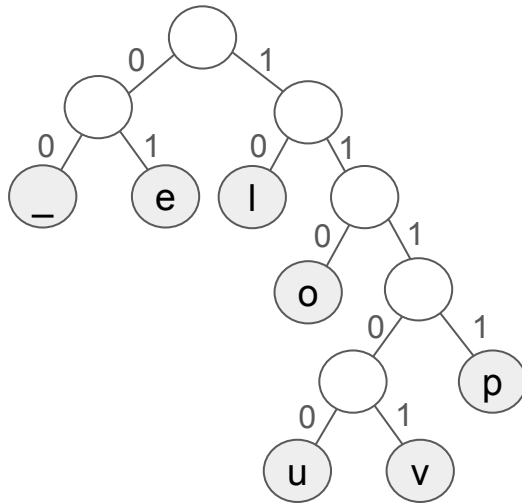


Codage de Huffman



Codage de Huffman

L'arbre se lit comme un dictionnaire : les valeurs sont des caractères et leur clé est leur codage.



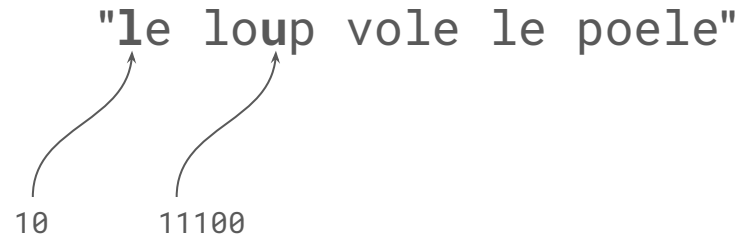
Caractère	Codage
l	10
e	01
-	00
o	110
p	1111
u	11100
v	11101

Codage de Huffman

Caractère	Codage
l	10
e	01
-	00
o	110
p	1111
u	11100
v	11101

"le loup vole le poele"

10 11100



Exemple de codage de deux des caractères

Codage de Huffman

Caractère	Codage Huffman	Codage naïf	Fréquence
l	10	000	5
e	01	001	5
-	00	010	4
o	110	011	3
p	1111	100	2
u	11100	101	1
v	11101	110	1

"le_loup_vole_le_poele" { 55 bits avec Huffman
63 bits avec codage naïf

(différence plus significative sur des données réalistes)