

Code ASCII

L'American Standard Code for Information Interchange (Code américain normalisé pour l'échange d'information, ASCII) est un codage simple très répandu en informatique. Les ordinateurs ne comprennent que le **langage binaire**, c'est-à-dire des suites de 0 et de 1. L'ASCII propose un dictionnaire entre notre langage et le langage binaire.

Alphabet ASCII				
-	11111		N	01110
A	00001		O	01111
B	00010		P	10000
C	00011		Q	10001
D	00100		R	10010
E	00101		S	10011
F	00110		T	10100
G	00111		U	10101
H	01000		V	10110
I	01001		W	10111
J	01010		X	11000
K	01011		Y	11001
L	01100		Z	11010
M	01101			

Lettre

Code ASCII

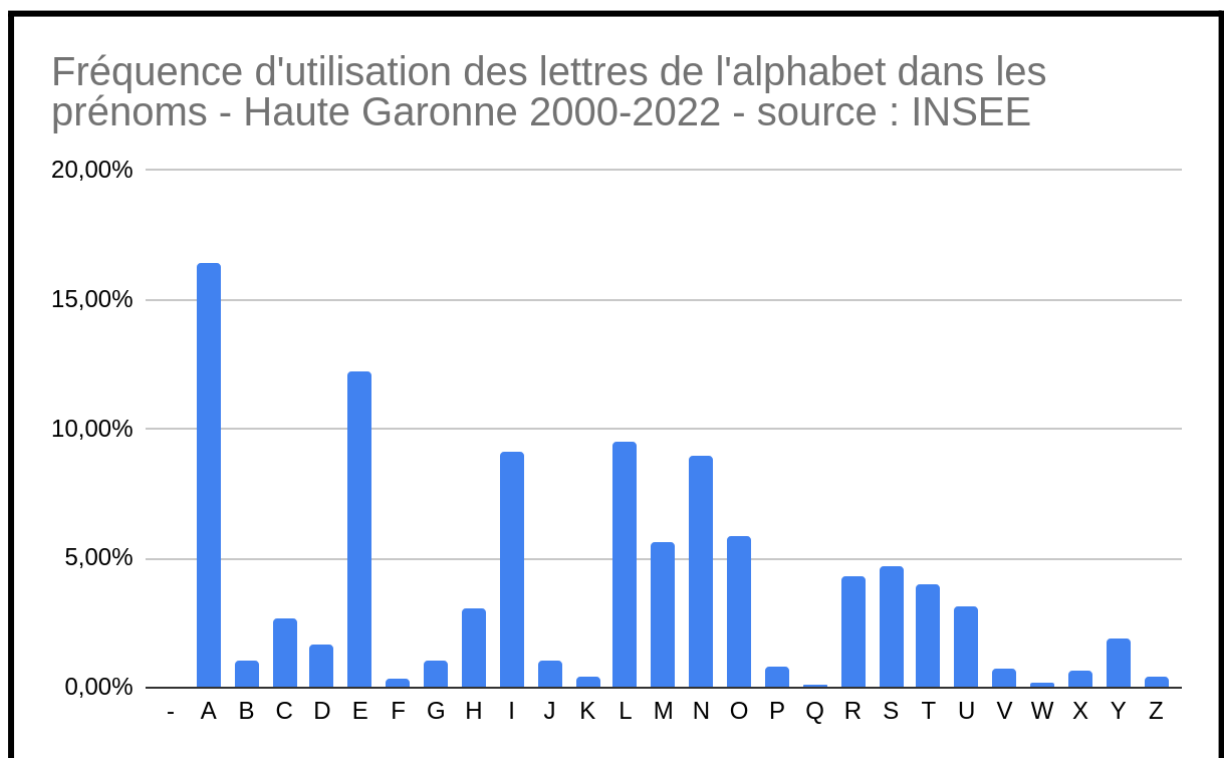
<hr/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<hr/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<hr/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<hr/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<hr/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<hr/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<hr/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<hr/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<hr/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<hr/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Codage de Huffman

Le codage de Huffman est une technique de **compression** sans perte de données basée sur les statistiques d'apparition des caractères dans le message.

Au lieu de coder chaque lettre sur 5 caractères (bits) on voudrait utiliser moins de caractères pour coder les lettres qu'on utilise souvent, pour que le message codé soit le plus court possible. Pour cela on se base sur la fréquence d'apparition des lettres dans le message à coder, ici les prénoms.

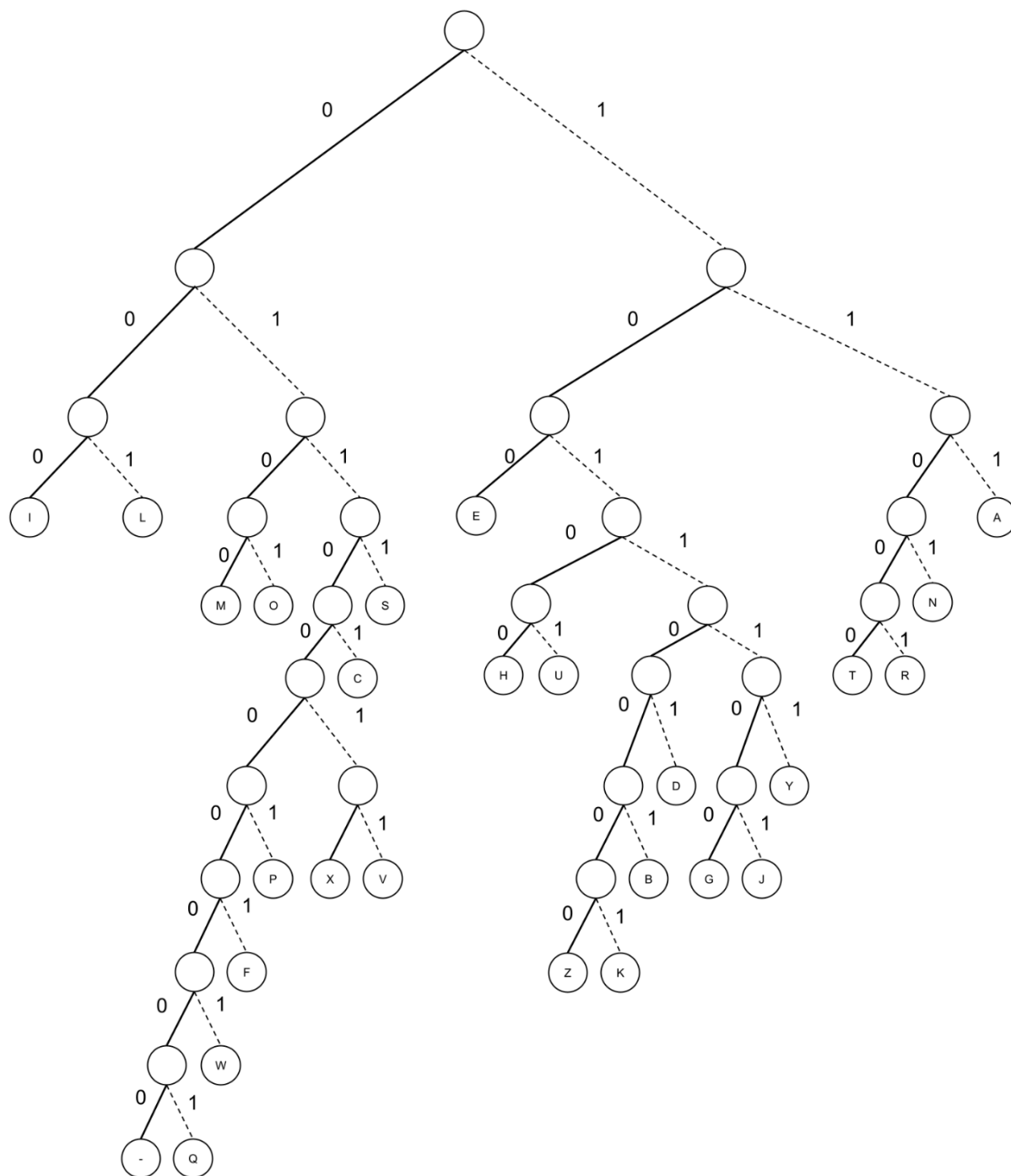
Ci-dessous la fréquence d'utilisation des lettres de l'alphabet dans les prénoms des personnes nées en Haute Garonne entre 2000 et 2022 (*source : Le Fichier des prénoms, INSEE*)



Intuitivement, on veut coder le "A" et le "E" avec un minimum de caractères, quitte à en utiliser trop dans les rares cas où un "Q" ou un "W" seraient utilisés.

Pour cela on va créer un **arbre de Huffman**, un réseau composé de nœuds et de liens, en groupant les lettres par deux en commençant par les plus rares. On ne s'étend pas ici sur la construction de l'arbre. Une fois créé, on affecte 0 aux liens de gauche et 1 aux liens de droite. Pour trouver le codage d'une lettre on part du haut (la racine) et on parcourt l'arbre jusqu'à la lettre (la feuille).

Arbre de Huffman des prénoms donnés en Haute Garonne entre 2000 et 2022
source:INSEE



Donc maintenant le “A” se code 111 tandis que le “Q” se code 0110000001.

En se basant sur ce nouveau codage, combien de caractères (bits) aurez vous besoin pour coder votre prénom ?

Alphabet Huffman			
–	01100000000	N	1101
A	111	O	0101
B	1011001	P	0110001
C	01101	Q	0110000001
D	101101	R	11001
E	100	S	0111
F	01100001	T	11000
G	1011100	U	10101
H	10100	V	0110011
I	000	W	011000001
J	1011101	X	0110010
K	10110001	Y	101111
L	001	Z	10110000
M	0100		