

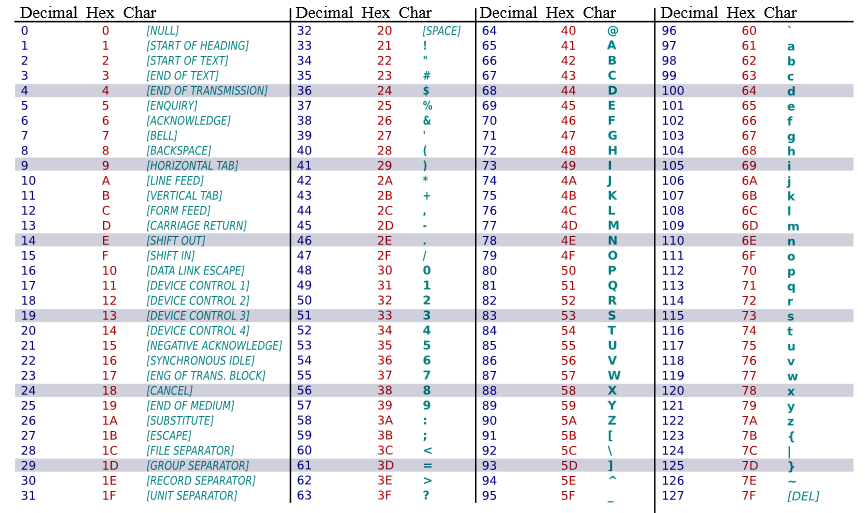
**V Représentation d’un texte en machine**

Nous avons appris à représenter les nombres, entiers ou flottants, en machine. Nous allons maintenant nous intéresser à la représentation des textes.

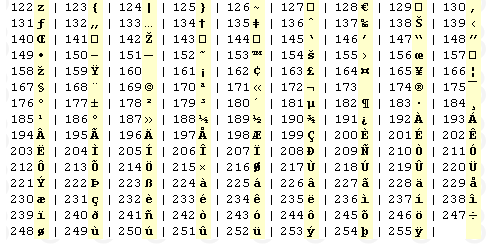
Contrairement au nombre, on est ici confronté au problème de l’harmonisation puisque il existe beaucoup d’alphabets différents, chaque langue ayant ses propres caractéristiques.

**1) Le code ASCII**

Le code ASCII a été conçu pour des textes écrits en anglais et permet de représenter 128 caractères, donc des mots de 7 bits. En fait, chaque caractère est codé sur un octet, le premier bit étant 0



L’inconvénient majeur ici est l’absence de codage possible pour les lettres accentuées. C’est pour cela qu’a été créée une extension à ce code, s’adaptant aux particularités des langues de l’Europe occidentale, même si le ‘œ’ a été oublié !!!!



On obtient ainsi une table quasi complète qui correspond à la norme **ISO-8859-1,** aussi appelée *latin-1*

**2)Unicode**

Cependant le problème de la concordance entre différentes langues n’était pas réglé. Avec le développement des communications mondiales entre machine, il fallut trouver un codage capable de faire coexister des textes européens, chinois, russe…

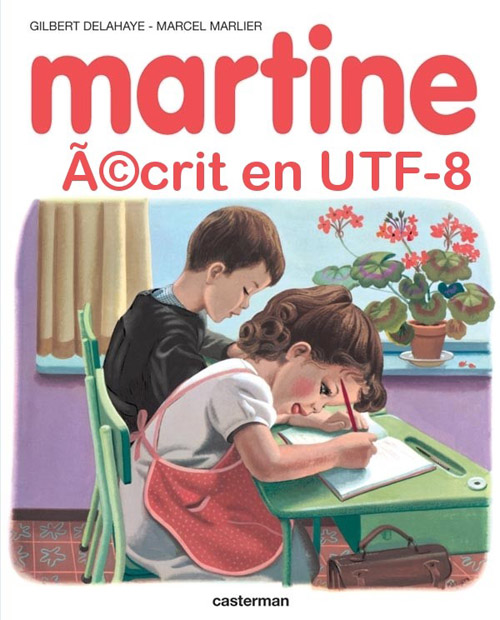
Un format universel a émergé, Unicode qui contient près de 110000 caractères. Il faut alors 32 bits pour coder un caractère !

Cependant, Unicode est juste une table de codage et ne s’occupe pas de la façon dont la machine encode les caractères. La longueur d’encodage est variable suivant les normes utilisées. Ainsi avec la norme **UTF-8** , les caractères les plus courants sont codés sur un octet, de façon compatible avec le code ASCII .

Voici la représentation binaire suivant le nombre d’octets utilisés



Si l’encodage UTF 8 est compatible avec la norme ASCII, il ne l’est pas avec la norme ISO-8859-1



**3)Exercices**

1) Coder vote prénom (en commençant avec une majuscule) à l’aide du code ASCII.

Quels sont selon vous les avantages et les inconvénients de ce codage ?

2) Trouvez le texte représenté en ASCII binaire par la suite de bits

010011100010111001010011001011100100100100101110

3) Ecrire un programme en langage python qui permet de coder un mot par la méthode de Vigenère.

<https://fr.wikipedia.org/wiki/Chiffre_de_Vigenère>

4) Ouvrez le site <https://apps.timwhitlock.info/unicode/inspect> et entrez cette phrase

Maître Corbeau, sur un 🌳 perché,

Précisez pour chaque caractère sur combien d’octet il est codé et donnez la représentation binaire du mot Maître.