

Codage des caractères : à savoir

Il existe différentes tables de codage des caractères.

Tout d'abord la table..... qui compte caractères. On code ceci surbits. En fait on utilise octet avec 1 bit servant à

Dans cette table il manque de la langue française et bien d'autres caractères des langues du monde entier.

On a donc créé des tables avec des compléments en occupant le dernier bit non utilisé.

On a en particulier la table ditequi compte caractères que l'on code sur

Cette table est entièrement avec la table ASCII.

Dans le but d'..... tous les codages un consortium a créé Cette norme attribue à chaque caractère un appeléet permet d'encoder tous les caractères existants.

Cette table est entièrement compatible avec la table mais seulement partiellement avec la table

Attention l'..... n'est pas un encodage des caractères.

Pour encoder l'..... on utilise souvent l'.....

Cet encodage n'utilise pas le même nombre pour chaque caractère.

On utilise 1 octet pour les caractères les plus fréquents (ceux de la table).

Pour les autres caractères on utiliseoctets.

C'est pour cela que par exemple "mathÃ©matique" apparaît avec deux caractères à la place du "é" quand le codage est mal détecté.

Pour reconnaître combien d'octets doivent être lus pour un caractère on regarde les

Exercice : Vérifier que le codage binaire de "Ã©" en ANSI correspond bien au codage UTF-8 de "é".

Ã correspond dans la table ANSI à xC3

© correspond dans la table ANSI à xA9

xC3A9 donne en binaire: b.....

On enlève les bits de poids forts qui servent à reconnaître le nombre

On obtient : b.....soit en décimal..... ce qui correspond bien à "é" !

En python (pour vérifier vos calculs) :

pour convertir un binaire (ex : 101010) en décimal : `int('101010',2)`

pour convertir un hexa (ex : C3) en décimal : `int('C3',16)`

pour convertir un décimal (ex : 195) en binaire : `hex(195)`

pour convertir un hexa (ex : C3 qui apparaît sous la forme 0xC3 dans python , c'est un zéro devant) en binaire : `bin(0xC3)` et `hex(0b.....)` dans l'autre sens.