

Activité : Encodage du texte

Les ordinateurs ne manipulent que des valeurs binaires, à savoir des 0 et des 1. Il a donc été nécessaire de traduire les textes en binaire afin d'envoyer des mails, rédiger des documents ou créer des pages web, etc.

Selon le lieu géographique et la langue parlée, les caractères utilisés doivent être transformés en un codage binaire. Les codages des caractères ont beaucoup évolué avec le temps et ont été regroupés sous forme de tables.

Nous allons nous intéresser à trois tables qui ont marqué l'évolution de l'encodage des caractères.

Avec un éditeur de texte, vous allez réaliser une présentation des trois tables de caractères en suivant le plan donné ci-dessous.

1) La première table de caractères est ASCII.

- a Que signifie l'acronyme ASCII ? Quelle est son origine ? Combien de caractères contient-elle ?
- b Insérer dans votre document une image de la table ASCII.
- c Quel est le nombre de bits utilisés pour chaque caractère ? Donner une plage (intervalle) du premier au dernier caractère ASCII.
- d Á quoi correspondent les 32 premiers caractères de cette table ? Et le dernier.
- e Quels caractères correspondent aux codes hexadécimaux 6E, 53 et 49 ? Donner pour chaque caractère sa valeur binaire et décimale.
- f Quel est le principal problème de cette table de caractères pour notre langue ?

2) Une seconde table introduite par l'ISO.

- a Que signifie l'acronyme ISO ?
- b Combien de tables contient la norme 8859 ?
- c Combien de caractères peut contenir une table à la norme ISO-8859 ?
- d Quel est le nombre de bits utilisés par chaque caractère ? Donner une plage (intervalle) du premier au dernier caractère ISO-8859 non ASCII.
- e Quelles ont été les deux table ISO-8859 utilisées en France ? Pourquoi ?

3) La norme Unicode

- a Quel est l'avantage de la norme Unicode sur la norme ISO ?
- b Qu'est-ce qu'un point de code dans la norme Unicode ? Comment se note-t-il ?
- c Quelles sont les points de code des caractères ASCII ?
- d Quel est le point de code associé au caractère È (alt+212) ? æ (alt+145) ? £ ?
- e Quel est le caractère associé au point de code est U+20AC ?
- f UTF-8 encode en binaire les caractères à partir de leur point de code. Combien d'octets sont nécessaires pour l'encodage d'un caractère ?
- g Donner selon les plages de valeurs Unicode l'encodage d'un caractère en UTF-8.