Activité : Comment sont codés les textes

Présentation

Tu as vu qu'un ordinateur manipule des informations binaires mais qu'il est plus facile de les manipuler des octets c'est à dire des **regroupements** de 8 bits. La notation hexadécimale est également utilisée pour avoir une écriture plus concise (plus courte).

Pour pouvoir **représenter** des textes en informatique, c'est à dire une suite de caractères, il est alors nécessaire de choisir d'associer à **chaque lettre un code binaire**.

Le codage des caractères : la table ASCII

Le **codage ASCII** est l'une des premières méthodes de codage des caractères: à chaque caractère est associé un code binaire sur un **octet**. Mais sur les 8 bits seuls 7 sont utilisés pour coder les caractères, le 8e bit, appelé **bit de parité**, est utilisé pour détecter les erreurs.

L'ensemble des caractères gérés par le codage ASCII est appelé le **jeu de caractères**.

Question : à partir de tes recherches sur Internet réponds aux questions suivantes :

- Que **signifie** le signe ASCII ?
- **Combien** de caractères est-il possible de gérer en ASCII ?
- Quels sont les **codes décimaux et hexadécimaux** des 3 caractères suivants : 4 S b ?
- Quels sont les caractères qui ont les **codes décimaux** : 34 40 123 ?
- Est-ce que la table ASCII de base permet de gérer les **caractères accentués** (**é**, **à**, ...) ?

Le codage des caractères : la table ASCII étendue

Pour gérer **davantage de caractères** il a été ensuite décidé d'utiliser le **8ème bit** pour ajouter au jeu de caractères ASCII du début, un jeu **étendu** de caractères. De cette manière, de nombreux jeux étendus ont vu le jour, pour coder des langues comme le **grec** ou le **russe** (alphabet cyrillique), mais aussi pour coder les langues d'Europe occidentale, en intégrant **accents et signes de ponctuations** qui nous sont propres.

L'Organisation internationale de normalisation (ISO) ainsi normalisé le jeu de caractère ISO-8859-1, ou ISO-Latin1 pour coder les langues d'Europe occidentale.

Question : à partir du site http://www.asciitable.com/ et de tes recherches sur Internet réponds aux questions suivantes :

- **Combien** de caractères est-il possible de gérer avec la table ASCII étendue ?
- Quels sont les **codes décimaux et hexadécimaux** des 3 caractères suivants : é à ?
- Lance un logiciel de traitement de texte comme Notepad ou LibreOffice. Tu peux écrire des caractères en saisissant leur **code ASCII en utilisant la touche ALT**. Quels sont les codes décimaux qui te permettent d'obtenir le dessin suivant :

La rédaction de documents en plusieurs langues

Les 256 caractères du jeu de caractères étendu ASCII ou ISO-8859-1 ne permettent pas de gérer des documents qui utilisent **plusieurs langues non européennes** comme le japonnais. Il est donc nécessaire de coder les caractères sur plusieurs octets tout en gardant la comptabilité avec la table ASCII. C'est ce que propose le jeu de caractère l'**UTF-8** (Unicode Transformation Format). On utilise ainsi entre 1 et 4 octets pour représenter un caractère, en introduisant une convention pour différencier les octets isolés,

A ce jour plus de de 90000 caractères sont actuellement gérés.

Question : à partir du site https://unicode-table.com/fr/ et de tes recherches sur Internet réponds aux questions suivantes :

- Que permet le codage **Unicode** ?
- **Combien** d'octets sont utilisés coder les caractères en Unicode ?

Les activités ...

Je reviens à la liste des activités.

From:
/ - **Les cours du BTS SIO**

Permanent link:
/doku.php/icn/facultatif/codagetexte

Last update: **2018/09/27 20:26**