

Un ordinateur ne manipule que des 0 et des 1 : Comment alors code-t-il du texte ?

I – Le code ASCII

1. Une première approche

Ouvrir un fichier Excell

- Sur la ligne 1 du tableau , écrire votre NOM et PRENOM en **MAJUSCULE**

-une case par lettre

- un espace entre NOM et PRENOM.

- Sur la ligne 2 du tableau , entrer la formule : = code (A1) puis recopier vers la droite :

A	B	C	D	E	F	G	H	I	J
N	O	M		P	R	E	N	O	M
78	79	77	32	80	82	69	78	79	77

Chaque lettre ainsi que l'espace est codée par un nombre selon le **code ASCII**.

- Même question mais avec des lettres minuscules . Qu'observez-vous par rapport au 1^{er} test ?

Culture Informatique...

*Un des premiers codages standardisés est le code du « Morse » au XIXe siècle.
Pour ce qui est de l'informatique , c'est le code ASCII dont les premières versions datent du début des années 1960.*

ASCII signifie **American Standard Code for Information Interchange**.

Le principe est assez simple, il consiste à associer à chaque lettre ou chiffre un entier entre 0 et 127.

Ce codage doit être compris par tous: il est normalisé.

*L'**ISO** (International Organization for Standardisation) est un organisme de normalisation international et l'**AFNOR** (Association française de normalisation) est l'organisme français.*

*Chaque caractère est identifié par un code unique qui est un entier naturel et la correspondance entre le caractère et son code est appelée un **Charset**.*

*Le code n'étant pas utilisable tel quel par un ordinateur qui ne comprend que le binaire, il faut donc représenter les codes par des octets: C'est l'**ENCODING**.*

Exercice : Petit calcul...

Sachant que l'on dispose de 127 caractères en ASCII , combien de bits sont nécessaires pour coder tous les caractères du code ?

- Utiliser la fonction **=DECBIN()** du tableur EXCELL pour obtenir la représentation en binaire de votre nom et prénom . Vérifier le nombre de bits utilisés pour chaque lettre .

A savoir....

En général , on code sur un octet , le premier bit étant toujours 0

2. Taille d'un texte

Exercice :

Quelle est la taille en octet de la phrase suivante :

(compter les lettres , les symboles et les espaces !)

«Enfin ! Je viens de comprendre ce qui se produit.»

Vérification :

- a. Taper cette phrase dans le Bloc notes de Windows ;
- b. Enregistrer ce texte en tant que « texte brut » (extension .txt)
- c. Vérifier la taille en octets du fichier (clic droit sur l'icône du fichier puis sur Propriétés)

Ecrivez la même phrase dans un logiciel de traitement de texte (Open office ou Word) et observez la taille du fichier . Quelle peut-être l'explication ?

3. Utilisation de la table ASCII (voir annexe 1)

- a. Quel est le code (en ASCII puis en binaire) du caractère 1 ? du caractère * ? (avec Excell)

- b. Coder en binaire : « L'an qui vient ! » puis vérifier avec Excell :

- c. Voici maintenant une exclamation codée en binaire :

01000010 01110010 01100001 01110110 01101111 00100001

Retrouvez cette exclamation :

Vérifiez avec Excel (on pourra utiliser la fonction CAR())

- d. Entrer la phrase suivante sous Excell et codez-là en ASCII :« Un âne est-il passé par là ?»
Qu'observez-vous ?

II – Quand la table ASCII ne suffit plus :

La norme ASCII permet de représenter un texte anglais ou un programme informatique, mais elle n'est pas adaptée pour les autres langues.

La table ASCII a été étendue pour pouvoir coder de nouveaux caractères. Le codage des caractères a été étendu sur 8 bits.

Ainsi, on peut citer la norme ISO 8859-1, définie dans les années 1980, aussi appelée « latin-1 » ou Europe occidentale, qui permet de représenter les caractères accentués (é, è, ç, à, ù, ô) mais omet quelques caractères fort utiles (ainsi, la ligature **oe** n'y figure pas).

Dans les pays occidentaux, cette norme a donné lieu à quelques extensions et adaptations, dont Windows-1252 et ISO 8859-15 (qui prennent en compte le symbole **€** créé après la norme ISO 8859-1).

La distinction entre ASCII, ISO 8859-1, ISO 8859-15 et Windows-1252 est une source de confusion parmi les développeurs de programmes informatiques car un même caractère peut être codé différemment suivant la norme utilisée. Par exemple, le caractère **€** a pour code **128** dans l'extension Windows-1252 et **164** dans la norme ISO 8859-15.

Sources :

http://fr.wikipedia.org/wiki/ISO_8859-1

<http://fr.wikipedia.org/wiki/Windows-1252>

http://fr.wikipedia.org/wiki/ISO_8859-15

Activité:

1. Combien le fait d'avoir 8 bits amène-t-il de nouvelles possibilités de codage de caractères?
2. Utilisation d'un convertisseur :

Se connecter au site : <http://nickciske.com/tools/binary.php>

Voici le code binaire d'un texte (que vous retrouverez sur votre bureau, nommé *binaire.txt*)

```
01000010 01110010 01100001 01110110 01101111 00101100 00100000 01110100 01110101
0010000001100001 01110011 00100000 01110000 01110010 01100101 01110011 01110001
01110101 0110010100100000 01110100 01101111 01110101 01110100 00100000 01110100
01110010 01101111 0111010101110110 11101001 00101110 00101110 00101110
```

→ À l'aide du logiciel fourni sur le site, retrouver le texte contenu dans le code.

→ Ce logiciel est-il compatible avec la norme ISO 8859-1, ISO 8859-15 ou Windows-1252 (ANSI) ? (justifier la réponse)

→ Trouver une astuce pour savoir laquelle des trois est utilisée !

III – Quand le net s'affole...

Nous avons tous un jour reçu un courriel bizarre ou lu une page web telle que celle-ci :

Prenons l'exemple typique de la lumière mise par un phare maritime : elle est d'abord indivisible, son coût de production est alors indépendant du nombre d'utilisateurs ; elle possède une propriété de non-rivalité (elle ne se détruit pas dans l'usage et peut donc être adoptée par un nombre illimité d'utilisateurs) ; elle est également non excluible car il est impossible d'exclure de l'usage un utilisateur, même si ce dernier ne contribue pas à son financement

Comment l'expliquer ?

IV – L'unicode

La globalisation des échanges culturels et économiques conduit à la coexistence de nombreuses langues : les langues européennes et de nombreuses autres langues aux alphabets spécifiques voire sans alphabet.

L'emploi d'internet dans le monde entier a donc nécessité la prise en compte d'un nombre plus important de caractères (par exemple, le mandarin possède plus de 5000 caractères).

D'autre part, avec le faible nombre de caractères pris en compte, cela pouvait conduire à des confusions : les symboles monétaires ne sont pas tous représentés dans le système ISO 8859-1 d'où des incompréhensions préjudiciables dans les ordres de paiement par courrier électronique.

Un consortium composé d'informaticiens, de chercheurs, de linguistes et de personnalités représentant les Etats ainsi que les entreprises s'occupe d'unifier toutes les pratiques en un seul et même système : la norme **UNICODE**.

L'Unicode est la table de correspondance Caractères code (**Charset**). ☒

L'UTF-8 est l'encodage correspondant (**Encoding**) le plus répandu (par défaut, les navigateurs Internet utilisent le codage UTF-8), il permet de coder tout caractère indépendamment de tout système de programmation ou système d'exploitation. Ce code est utilisé dans les pages WEB, les emails, etc...

Le codage UTF-8 est compatible avec la norme ASCII : tout caractère ASCII se code en UTF-8 sous forme d'un unique octet, identique au code ASCII.

Le codage UTF-8 a vocation à devenir le standard, mais il ne l'est pas encore, malgré les efforts des comités de normalisation!

Par défaut, les navigateurs Internet utilisent le codage UTF-8 et les concepteurs de sites pensent de plus en plus à créer leurs pages web en prenant en compte cette même norme; c'est pourquoi il y a de moins en moins de problèmes de compatibilité.

Activité :

Ouvrir la page web du Lycee du Grand Noumea avec Firefox ou internet Explorer

Rechercher le Code Source de la page en cliquant sur « **Affichage** » puis « **Encodage des caractères** »

Quel est le code choisi ?

Changer le code et observer l'écriture des caractères.