

<p style="text-align: center;">info223 : Science informatique TD 2 : représentation des caractères, dates et autres</p>

Pierre Hyvernât
Laboratoire de mathématiques de l'université de Savoie
bâtiment Chablais, bureau 22, poste : 94 22
email : Pierre.Hyvernât@univ-savoie.fr
www : <http://www.lama.univ-savoie.fr/~hyvernât/>

Partie 1 : Caractères

Rappels :

- ISO-8859-*n* sont des extensions de l'ASCII pour utiliser le bit de poids fort (sur un octet),
- UTF-8 (UCS Transformation Format 8 bit, UCS = Universal Character Set) est un encodage de l'Unicode (sur un, deux, trois ou quatre octets),
- UTF-16 est un autre encodage de l'Unicode (sur deux ou quatre octets).

Question 1. Est-il impossible, difficile, facile de transformer un fichier texte :

- de l'ASCII vers l'ISO-8859-*n*
- de l'ASCII vers l'UTF-8
- de l'ASCII vers l'UTF-16
- de l'ISO-8859-*n* vers l'ASCII
- de l'ISO-8859-*n* vers l'ISO-8859-*m*
- de l'ISO-8859-1 vers l'UTF-8
- de l'UTF-8 vers l'ISO-8859-1
- de l'UTF-8 vers l'UTF-16
- ...

Que pouvez-vous faire lorsque la traduction est impossible ?

Question 2. Avec l'ASCII ou les ISO-8859-*n*, la taille en octets d'une chaîne de caractères est égale à son nombre de caractères. Ce n'est plus le cas pour l'UTF-8.

Pour chacune des chaînes ci-dessous, dites si la chaîne (représentée en hexadécimal) est de l'UTF-8 valide, et si oui, combien de caractères elle contient :

- 63 6f 75 63 6f 75
- 48 e9 20 21
- 48 c3 a9 20 21

Question 3. Dans certains langages de programmation, vous pourrez voir des opérations sur les caractères comme :

- `if (c1 < c2) ...`
- ou `c = c-32`

Que se passe-t'il, et qu'en pensez-vous ?

Partie 2 : Dates, images

Question 1. Calculez la date correspondant à 4d 4e 76 d9 si la représentation est celle du système de fichier FAT (date sur les octets de poids fort, heure sur les octets de poids faible),

Question 2. Même question s'il s'agit d'une date Unix...

Question 3. Donnez une manière simple de récupérer l'année d'une date au format FAT si elle vous est donnée sur un entier 32 bits.

Question 4. Écrivez un petit programme en Python qui crée une image “.ppm” carrée de taille 32 pixels par 32 pixels avec :

- un fond noir,
- une diagonale rouge.

Partie 3 : Compression des données

Question 1. Est-il possible qu’un programme de compression (`gzip`, `bzip2`, `zip` ou autre) puisse compresser tous fichiers ?

Question 2. L’ADN est formé de 4 bases : l’adénine, la thymine, la cytosine et la guanine. Comment stocker efficacement des séquences d’ADN dans la mémoire ?

Question 3. Si l’on veut stocker des suites des caractères n’utilisant que les voyelles, comment peut-on procéder pour ne pas perdre trop de place.

Question 4. On souhaite coder un texte en français pour optimiser l’utilisation de la mémoire. On veut par exemple utiliser le fait que la lettre “e” est beaucoup plus présente.

Discutez et proposez des solutions...