

Encodages de caractères : Exercices sur feuille

Exercice A :

Encoder en ASCII la chaîne de caractères suivante.

"Ada Lovelace"

On représentera les caractères encodés sous forme binaire par codets de 7 bits.

En hexadécimal :

41 64 61 20 4c 6f 76 65 6c 61 63 65

En binaire par codets de 7 bits:

1000001 1100100 1100001 0100000 1001100 1101111 1110110 1100101 1101100 1100001
1100011 1100101

En binaire par codets de 8 bits (par octets) :

01000001 01100100 01100001 00100000 01001100 01101111 01110110 01100101 01101100
01100001 01100011 01100101

Exercice B :

Décoder la séquence suivante de caractères encodés en ISO-8859-1, représentée en hexadécimal :

0x50 0xE8 0x72 0x65 0x20 0x4E 0x6F 0xEB 0x6C

On obtient 'Père Noël'

Exercice C :

Encoder en UTF-8 la chaîne de caractères suivante :

"Ôu ça ?"

On représentera les caractères encodés sous forme binaire.

Ô : pt de code U+00D2 <-> 11010010 entre 8 et 11 bits : 11000011 10010010

u : pt de code U+0075 <-> 01110101 moins de 7 bits : 01110101

: pt de code U+0020 <-> 00100000 moins de 7 bits : 00100000

ç : pt de code U+00E7 <-> 11100111 entre 8 et 11 bits : 11000011 10100111

a : pt de code U+0061 <-> 01100001 moins de 7 bits : 01100001

: pt de code U+0020 <-> 00100000 moins de 7 bits : 00100000

? : pt de code U+003F <-> 00111111 moins de 7 bits : 00111111

Ce qui donne finalement :

11000011 10010010 01110101 00100000 11000011 10100111 01100001 00100000 00111111

Exercice D :

Décoder la séquence suivante de caractères encodés en UTF-8, représentée en binaire :

11000011 10100000 00100000 01100011 11000101 10010011 01110101 01110010 00101110

11000011 10100000 00100000 01100011 11000101 10010011 01110101 01110010 00101110

11000011 10100000 | 00100000 | 01100011 | 11000101 10010011 | 01110101 | 01110010 |
00101110

00011100000 | 0100000 | 1100011 | 00101010011 | 1110101 | 1110010 | 0101110

U+00E0 | U+0020 | U+0063 | U+0153 | U+0075 | U+0072 | U+002E

à cœur.

Encodages de caractères : Exercices sur machine

Exercice 1 :

1. Ouvrir avec un navigateur le fichier mon_expose.html situé dans ./ressources/html_et_encodage/.
2. Que constate-t-on ?
3. Ouvrir ce même fichier avec l'éditeur de texte Notepad++ :
 - a) En bas à droite de la fenêtre, s'affiche l'encodage utilisé. Quel est cet encodage utilisé ?
 - b) Dans l'en-tête du fichier, quel est le format d'encodage indiqué ?
 - c) Corriger le problème en indiquant le format d'encodage "UTF-8".
 - d) Vérifier l'efficacité de la solution.
4. On peut inclure n'importe quel caractère Unicode dans un document html encodé en UTF-8 en utilisant la syntaxe suivante :

&#N;

En remplaçant N par le point de code Unicode en *décimal*.

Ou en utilisant la syntaxe suivante :

&#xN;

En remplaçant N par le point de code Unicode en *hexadécimal*.

- a) Tester cela en introduisant des caractères UTF-8 "spéciaux" indiqués dans les plans Unicode fournis dans le cours dans le document html.
- b) Un des six plans de code fournis a des caractères que le navigateur n'est pas capable d'afficher. De quel plan de code s'agit-il ?
- c) Pourquoi le navigateur n'arrive-t-il pas à afficher certains caractères ?