

TP5

La page web se trouve ici :

<http://www.lsv.ens-cachan.fr/~schwoon/enseignement/systemes/ws1415/>.

1 Codage des caractères en HTML

Les pages HTML connaissent plusieurs façons de coder des caractères :

- Les *références d'entités* qui permettent de décrire des caractères non-ASCII par des caractères ASCII, et ça indépendamment du codage utilisé dans le reste du fichier. P.ex., à s'écrit `à` dans cette notation.
- Les *références numériques* qui donnent le code d'un caractère Unicode, dans la forme `&#nnn;` (décimal) ou `&#xnnn;`.
- Le codage direct binaire dans l'un des formats qu'on a discuté dans le cours (UTF-8, ISO-8859-1, ...). Ceci nécessite de rajouter une balise dans la forme

```
<meta http-equiv="Content-Type" content="text/html;  
charset=iso-8859-1" />
```

dans la partie `head` du document HTML.

On va considérer les noms des communes européennes suivantes :

Crèvecoeur (France)
L'Haÿ-les-Roses (France)
Kroměříž (Tchéquie)
Gödöllő (Hongrie)
Præstø (Danemark)

1. Créez un document HTML qui affiche ces noms avec leurs références d'entités.
2. Cherchez les codes Unicode des caractères non-ASCII et utilisez-les pour recréer la liste avec des références numériques.
3. Créez un document HTML en UTF-8 direct avec votre éditeur de texte.

2 UTF-8

1. La page web du cours contient un programme C incomplet pour générer les codets UTF-8 à partir d'un entier. Complétez la fonction `utf8`. Il suffit de remplir les éléments de `buf`, en utilisant les bonnes opérations binaires (`&`, `>>`) sur la valeur `cp`.
2. Créez un programme qui fait l'inverse : décoder un flux d'octets codant un texte en UTF-8 et émettre les codes Unicode qui composent le texte. Afficher une erreur si jamais le flux contient un codage incorrect.

3 Formater du texte

La page web contient deux programmes qui affichent les noms de quelques villes et leur population. Dans l'un des programmes les noms sont codés en iso-8859-1, dans l'autre en utf-8.

1. Compilez et exécutez les deux programmes. L'affichage, est-elle bon pour les deux ? Pourquoi pas ?
2. Trouvez une méthode pour corriger l'affichage du programme qui ne fonctionne pas.