

Documents numériques  
Majeure Informatique  
Exercices 2017–18

Michel Beigbeder

### 3 Exercices de programmation en sh, 10 octobre 2017

#### Exercise6 TP : Codage et encodage des caractères

- 6.1. Créez avec un éditeur de texte un fichier `petitTexte.txt` avec le contenu :
- 01  
Bien sûr il doit s'agir d'un fichier de texte brut (*plain text* en anglais).  
Autrement dit votre fichier doit contenir la séquence de caractères :
- DIGIT ZERO  
DIGIT ONE  
LINE FEED (LF)
- Examinez en hexadécimal le contenu de ce fichier.
- 6.2. Modifiez le fichier `petitTexte.txt` pour qu'il contienne maintenant :
- âéêè  
Autrement dit votre fichier doit contenir la séquence de caractères :
- LATIN SMALL LETTER A WITH GRAVE  
LATIN SMALL LETTER E WITH ACUTE  
LATIN SMALL LETTER E WITH CIRCUMFLEX  
LATIN SMALL LETTER E WITH GRAVE  
LINE FEED (LF)
- 6.3. Quel est le nombre de caractères de ce fichier ?
- 6.4. Quel est le nombre d'octets de ce fichier ?

**Remarque :** *Il est probable que votre fichier soit encodé en UTF8 ou en ISO-8859-1, mais ça pourrait être autre chose et de toutes façons cela dépend de votre environnement de travail, de l'éditeur de texte que vous avez utilisé, toutes choses que je ne contrôle pas.*

*Pour convertir un fichier dans un autre jeu de caractères vous pouvez utiliser soit votre éditeur de texte s'il en est capable soit l'un ou l'autre des utilitaires `iconv` ou `uconv` selon ce qui est installé sur votre machine.*

ICONV(1)

Linux User Manual

ICONV(1)

NAME

`iconv` - convert text from one character encoding to another

UCONV(1)

ICU 52.1 Manual

UCONV(1)

NAME

`uconv` - convert data from one encoding to another

- 6.5. Encodez le contenu de ce fichier en ISO-8859-1 en mettant le résultat dans `petitTexte-latin1.txt` ? Quel est le nombre de caractères de ce nouveau fichier ? Quel est le nombre d'octets de ce nouveau fichier ? Quels sont les numéros des caractères de ce fichier en ISO-8859-1 ? Est-ce que votre éditeur de texte permet de charger « correctement » ce fichier ?
- 6.6. Encodez le contenu de ce fichier en UTF8 en mettant le résultat dans `petitTexte-utf8.txt` ? Quel est le nombre de caractères de ce nouveau fichier ? Quel est le nombre d'octets de ce nouveau fichier ? Quels sont les numéros des caractères de ce fichier en Unicode ? Quels sont les encodages en UTF8 de ces caractères ? Est-ce que votre éditeur de texte permet de charger « correctement » ce fichier ? À partir du numéro en ISO-8859-1 de ces caractères, construisez « à la main » leur encodage en UTF8, et si vous ne retrouvez pas ce que vous avez extrait du fichier, cherchez l'erreur.
- 6.7. Créez un fichier avec votre navigateur de fichier dont le nom est `é.txt` ? Dans quel jeu de caractères et quel encodage est ce nom de fichier ?
- 6.8. Insérez le caractère `œ` (LATIN SMALL LIGATURE OE) dans votre petit fichier de texte. Pouvez-vous le sauvegarder en UTF8 ? Pouvez-vous le sauvegarder en ISO-8859-1 ?