

Travail dirigé - L'encodage de texte

1. Exercice 1 : Détermination de la taille d'un texte

Question 1) **Enregistrer** un fichier texte encodé au format ISO-8859-1 contenant uniquement la lettre « e » (ne pas appuyer sur Entrée après la lettre) puis un autre avec la lettre « é ». **Relever** la taille des deux fichiers. **Justifier** ce résultat.

Les deux fichiers contiennent 1 octet. C'est normal puisque cette norme est une évolution du code ASCII sur 8 bits. Tous les caractères sont codés sur 8 bits

Question 2) **Reprendre** la question 1 en enregistrant les fichiers au format UTF-8.

Caractère	Taille en UTF-8
e	1 octets
é	2 octets

L'encodage UTF-8 est un codage à longueur variable suivant les caractères encodés.

Question 3) Sous Windows, **enregistrer** un fichier texte encodé au format UTF-8 ou ISO-8859-1 contenant uniquement la lettre « e » et appuyer sur la touche Entrée afin de faire un retour à la ligne. **Relever** la taille du fichier. **Justifier** ce résultat.

La taille du fichier a augmenté de 2 octets quelque soit l'encodage utilisé. Sous windows le retour à la ligne est codé par les deux caractères CR+LF (\$0D et \$0A)

2. Exercice 2 : Etude de la taille d'un texte

On demande à des personnes d'écrire sur leur ordinateur en codage ISO-8859-1 le texte suivant : «**Enfin ! Je viens de comprendre ce qui s'est produit.**»

On demande ensuite aux participants la taille du fichier qu'ils ont écrit.

Remarque concernant les caractères spéciaux

Certains caractères sont spéciaux. Par exemple LF correspond au « \n » des programmeurs et permet historiquement de créer une nouvelle ligne. On peut également citer CR correspondant au retour chariot et marquant la fin d'une ligne « \r ». Sur les cartes de programmation (en effet, les écrans n'ont pas toujours été utilisés...) pour signifier un retour au début de la ligne il fallait faire une combinaison CR+LF.

Cette combinaison est toujours utilisée sous Windows pour faire un retour à la ligne, alors que ce n'est pas le cas sous Mac OSX et Linux.

Question 4) **Relier** par rapport à la taille fournie par les participants les conditions dans lesquels le code a été enregistré.

- | | | |
|--------------------|---|---|
| Taille : 49 octets | • | • La taille du fichier est la bonne. |
| Taille : 51 octets | • | • Vous avez mis beaucoup trop d'espaces non ? |
| Taille : 52 octets | • | • Vous êtes sous Windows et après avoir tapé la phrase vous avez appuyé sur la touche sur Entrée. |
| Taille : 53 octets | • | • Avez-vous bien tapé la phrase ? |
| Taille : 54 octets | • | • Vous avez oublié un espace ou le caractère « . ». |
| Taille : 56 octets | • | • Vous êtes sous Linux et après avoir tapé la phrase vous avez appuyé sur la touche sur Entrée. |

