

LABORATOIRE 10

Buts

Lire et écrire dans des fichiers

Exercer l'utilisation d'une liste chaînée

Donnée

Écrire un programme qui lit un fichier de texte (au format UTF8 simplifié), en extrait les mots, et génère un fichier les listant avec le nombre d'occurrence de chacun.

Indications :

Le format UTF8 permet de coder des caractères sur 1, 2, 3 ou 4 octets. Les lettres ont les même codes que l'ASCII. Les caractères codés sur plus d'un octets commencent par une balise, non utilisée pour coder un caractère sur 1 octet. On se restreindra aux balises `0xC3` et `0xC5`, qui permettent de coder sur 2 octets (en tout) les lettres spéciales du français, la balise `0xC2` qui code sur 2 octets des signes de ponctuation, et de la balise `0xE2` qui code sur 3 octets d'autres signes de ponctuation. À noter que `0x0D` et `0x0A` marquent le retour de chariot avec une nouvelle ligne.

Déclarer un type `LISTE` dont les éléments contiendront un entier pour compter le nombre d'occurrences et un pointeur vers un tableau de caractères ; la taille de ce tableau sera réservée dynamiquement en fonction du nombre de lettres du mot.

Écrire ensuite une fonction permettant d'introduire un nouveau mot ne figurant pas dans la liste, ou d'incrémenter le compteur d'occurrences d'un mot déjà dans la liste. Pour simplifier, on considérera que la casse est significative, de même que les signes diacritiques.

Considérer 2 versions de cette fonction : l'insertion en queue de liste d'un nouveau mot et une insertion maintenant l'ordre lexicographique des mots de la liste (prendre simplement le résultat de la fonction `strcmp` pour le classement).

Écrire ensuite une fonction `parcourir(liste, fichier)` permettant de sauver tous les éléments de la liste dans un fichier

Finalement, écrire une fonction `liberer(liste)` qui restitue intégralement la mémoire

Questions

Y a-t-il une différence de performance entre l'introduction d'un nouveau mot en queue de liste ou en maintenant l'ordre lexicographique ? Si oui, en expliquer la raison.

Dans le fichier `les_miserables.txt`, y a-t-il un ou des mots apparaissant plus souvent que « le » ?

Sur la dernière page, y a-t-il un mot n'apparaissant nulle part ailleurs dans le texte ?

Quel personnage est le plus cité ?

À rendre à la dernière séance de laboratoire. Le travail donnera lieu à une évaluation.