

CPP - TP1

L'objectif de ce TP est d'appliquer les principes de la structuration des données pour mettre en oeuvre un programme de traitement de fichiers. La taille importante des fichiers considérés impose que la conception des algorithmes des programmes demandés conduise au choix des structures les plus adaptées pour éviter d'en arriver à des temps d'exécution excessifs, voire rédhibitoires.

1 CRÉATION D'UN LEXIQUE

1.1 CARACTÉRISTIQUES D'UN LEXIQUE

Nous cherchons dans un premier temps à créer un lexique de tous les mots contenus dans un texte donné.

Vous créerez une classe `lexique` caractérisée par un nom et pour chaque mot du lexique, son nombre d'occurrences. Choisissez le conteneur qui vous semble le plus pertinent. Le lexique devra être constitué à partir d'un texte contenu dans un fichier d'entrée.

Vous avez à disposition deux textes "lesMiserables_A.txt" et "notreDameDeParis_A.txt" de Victor Hugo, en anglais pour minimiser les problèmes d'accents.

Remarque : Vous pourrez utiliser <https://www.cplusplus.com/reference/cstring/strtok/> pour découper une chaîne de caractères en fonction de délimiteurs. Il vous est fourni quelques fonctions utilitaires dans `utilitaire.cpp` et `utilitaire.h`.

1.2 OPÉRATION SUR UN LEXIQUE

Vous définirez à minima les opérations suivantes :

- sauvegarder le contenu du lexique ainsi produit dans un fichier de sortie;
- tester la présence d'un mot dans le lexique et retourner son nombre d'occurrences;
- supprimer un mot du lexique;

- afficher le nombre de mots différents présents dans le lexique.

Vous pouvez ajouter toutes les fonctions qui vous semblent pertinentes pour résoudre le problème.

Par surcharge d'opérateurs, vous permettrez de :

- `<<` : afficher le lexique à l'écran (opérateur externe) ;
- `+=` : fusionner deux lexiques (opérateur interne à la classe `Lexique`) ;
- `-=` : faire la différence de deux lexiques, ie. ne garder que les mots dans le premier lexique qui ne sont pas dans le deuxième (opérateur interne à la classe `Lexique`).

Précisez dans votre rapport le diagramme de classe de la classe `lexique`, les algorithmes de la fusion et de la différence de deux lexiques. Vous illustrerez le bon fonctionnement de vos programmes par des jeux d'essais.

2 CRÉATION D'UN LEXIQUE AVEC NUMÉRO DE LIGNE

Vous créerez ensuite un lexique qui héritera de votre première classe, `lexique` mais qui permet de garder en plus pour chaque mot du lexique la ligne du fichier où il a été rencontré. Vous pourrez utiliser <https://www.cplusplus.com/reference/string/string/getline/>. Vous êtes libre de choisir la conception que vous souhaitez, mais vous devez la détailler dans un diagramme de classes.

3 RENDUS

Merci de déposer vos solutions sur hippocampus (rapport en pdf et codes ou lien github).