

Indexation de texte

Exercice 1 - L'indexeur

Le but de ce TD est d'écrire un programme effectuant de l'indexation de texte.

Fichier TXT	Fichier IDX
ligne 1: hello ... tutu	hello: 1 2
ligne 2: ... hello	tutu: 1 3
ligne 3: tutu ...	

Un indexeur est un programme qui prend en entrée un fichier texte et qui crée l'index correspondant. Un index est un objet sur lequel on peut effectuer les requêtes suivantes :

- 1 Indexer un fichier donné ;
- 2 Obtenir la liste des mots distincts du texte ;
- 3 Obtenir la liste des occurrences d'un mot ;
- 4 Indiquer les occurrences des mots commençant par les lettres 'xxy' ;
- 5 Indiquer si un mot est présent ou non dans la liste ;
- 6 Charger/Sauver un index déjà existant.

Le fichier correspondant un index possèdera l'extention IDX. Il est composé de lignes correspondant à un mot donné et indiquant les occurrences sous forme du numéro de ligne du mot dans le fichier à indexer. Les occurrences seront triées par ordre croissant et les mots par ordre lexicographique.

```
java Indexeur fichier.txt           // crée le fichier.idx
                                     // correspondant
java Indexeur fichier.idx -list     // affiche les mots trouvés
java Indexeur fichier.idx -exist mot // indique si le mot existe
java Indexeur fichier.idx -occ mot  // affiche les occurrences
                                     // du mot trouvé
java Indexeur fichier.idx -start mot // affiche les occurrences
                                     // des mots commençant
                                     // par le mot mot
```