

Université Paris-Est Créteil
IUT de Sénart-Fontainebleau
Département informatique
IS3
Algorithmique Programmation
2014–2015

Partiel
Durée : 1 heure 15 mn

Seuls sont autorisés, à titre de documents, les listings comprenant explicitement le nom de l'étudiant (et celui-ci seulement) et les notes manuscrites.

Il arrive que l'on ait besoin de connaître les différents mots se trouvant dans un texte. Pour cela il est intéressant d'obtenir la liste de ces mots dans l'ordre alphabétique.

De nos jours, le texte sera un fichier (texte) et la liste obtenue sera placée dans un autre fichier texte, un mot par ligne.

On sait récupérer le texte du fichier texte ligne par ligne. Séparer les mots est tout un problème en soi. En Java, on peut utiliser un outil appelé « tokenisation ».

Le tri peut être effectué avec un arbre binaire de tri (de chaînes de caractères).

- 1^o Écrire une classe Java `StringTree` des arbres binaires dont les nœuds sont étiquetés par des chaînes de caractères.

[On n'oubliera pas la classe auxiliaire des nœuds, contenant elle-même une méthode d'insertion. La classe elle-même contiendra un constructeur, une méthode d'insertion d'une chaîne de caractères.]

- 2^o) Ajouter à la classe `StringTree` une méthode :

```
public void sort(PrintStream)
```

qui place dans la fichier passé en argument les étiquettes de l'arbre, dans l'ordre alphabétique.

[On aura intérêt à utiliser une méthode auxiliaire privée dont les arguments sont un nœud et un fichier.]

- 3^o) Écrire une application Java demandant le nom d'un fichier (texte) source, le nom d'un fichier (texte) but et qui place dans le fichier but la liste des mots se trouvant dans le texte contenu dans le fichier source dans l'ordre alphabétique, un mot par ligne, sans répétition.

Documentation

1 Chaîne de caractères

Rappelons que la méthode :

```
public int compareTo(String)
```

de la classe `String` renvoie un entier strictement positif lorsque la chaîne de caractères à laquelle on l'applique vient avant la chaîne de caractères argument dans l'ordre alphabétique (déduit du code Unicode).

2 Découpage en mots

Le paquetage `java.util` contient la classe :

```
StringTokenizer
```

permettant de découper une chaîne de caractères en mots. Elle possède :

- un constructeur dont l'argument est une chaîne de caractères (celle à découper en mots) ;

- une méthode :

```
public String nextToken()
```

renvoyant le mot suivant de la chaîne de caractères si elle en possède encore ;

- une méthode :

```
public boolean hasMoreTokens()
```

indiquant s'il reste des mots non encore récupérés dans la chaîne de caractères initiale.