

Rapport de projet d'algorithmique 2 : structures récursives, linéaires et binaires

KLAK Karolina et MASSIAS Antoine ¹

Janvier 2022

¹Rapport réalisé avec LaTeX

Contents

1	Généralités	2
1.1	Spécifications	2
1.2	Options	3

Chapter 1

Généralités

Ce projet constitue l'aboutissement du projet d'algorithmique 2, du semestre 3 de la licence d'informatique à l'UFR Sciences et Techniques de Rouen.

Le projet a été réalisé par A. Massias et K. Klak dans le cadre de leurs études. Ce projet consiste en la réalisation, en C, d'un exécutable produisant sur la sortie standard une liste de mots partagés. En effet, cet exécutable doit, au terme de plusieurs noms de fichiers inscrits sur la ligne de commande de trier les mots figurant dans chacun d'eux, et d'en conclure une liste de "mots" dits "partagés" entre les fichiers, tout en spécifiant leur motif et leur nombre total d'occurrences. A titre informatif, le motif fait référence aux fichiers dans lesquels ledit mot apparaît. Il se compose d'une suite de 'x' ou '-' en fonction de la présence ou non du mot dans chacun des fichiers.

1.1 Spécifications

La liste dont il est question est triée selon trois filtres :

- ordre décroissant du nombre de fichiers dans lesquels le mot figure,
- ordre décroissant des nombres totaux d'occurrences,
- ordre lexicographique croissant des mots.

Une ligne est réalisée pour chaque mot, avec le descriptif ci-dessus. Chaque liste est composée de dix lignes, limite fixée par défaut qu'il est possible de modifier par le biais d'options (cf. 1.2 Options). Une autre limitation fixée par défaut : le nombre de caractères de chacun des mots étant pris en compte. En effet, par défaut seuls les 63 premiers caractères d'un mot sont pris en compte. Au-delà de cette limite, le mot est coupé.

Exemple : L'exécutable produit par défaut les lignes de textes suivantes :

```
$ ./ws textes/tartuffe.txt textes/lesmiserables.txt
```

```
./ws: Word from 'textes/lesmiserables.txt' at line 33239 cut: 'paroisse-meurs-de-faim-si-tu-as-du-feu-meurs-de-froid-si-tu-as-...'
```

```
./ws: Word from 'textes/lesmiserables.txt' at line 62928 cut: '-L'agent-de-police-Ja-vert-a-ete-trouve-noye-sous-un-bateau-du-...'
```

```
xx 22021 de
```

```
xx 15256 la
```

```
xx 13062 et
```

```
xx 11685 a
```

```
xx 11194 le
```

```
xx 7252 les
```

```
xx 6114 un
```

```
xx 5667 que
```

```
xx 5444 il
```

```
xx 5375 qui
```

Il y a précision si des mots ont été coupés. Ainsi comme convenu seules 10 lignes de textes sont affichées.

Le motif ici est 'xx' pour chacun des mots de la liste : cela signifie que les mots apparaissent dans les deux fichiers.

Le nombre qui suit indique le nombre d'occurrences totales, dans tous les fichiers.

Enfin, en toute logique, la dernière chaîne de caractères de la ligne correspond au mot, entier ou coupé.

1.2 Options

En totalité, cinq options coexistent : deux d'entre elles concernent le nombre de lignes de mots, une concernant la longueur maximale du préfixe du mot, et enfin deux sur la notion du mot.

En outre :

- '-t' : limite du nombre de lignes de textes (limite du nombre de mots à produire)
- '-s' : production de lignes supplémentaires de tous les possibles mots dont le motif et le nombre d'occurrences sont égaux à ceux du dernier des mots associés à la limite
- '-i' : fixer la longueur maximale du préfixe de chaque mot
- '-p' : permet de ne plus considérer une marque de ponctuation comme un "mot"
- '-u' : permet de convertir un caractère correspondant à une lettre minuscule en un caractère de lettre majuscule

Exemple : ci-dessous les lignes de textes en sortie avec l'utilisation des 5 options.

```
$ ./ws -t 12 -s -i 5 -p -u textes/toujoursetjamais.txt textes/lecid.txt textes/lesmiserables.txt  
textes/lavare.txt
```

```
...
```

```
./ws: Word from 'textes/lecid.txt' at line 34 cut: 'TEMPS...'.  
./ws: Word from 'textes/lecid.txt' at line 34 cut: 'VAILL...'.  
xxxx 18020 LA  
xxxx 14920 ET  
xxxx 13995 A  
xxxx 13407 LE  
xxxx 11824 L  
xxxx 8742 UN  
xxxx 8396 LES  
xxxx 6531 D  
xxxx 6246 UNE  
xxxx 6169 EST  
xxxx 5943 QUI  
xxxx 5671 DANS
```