

Projet d'algorithmique du texte

La commande Sed

20 octobre 2016

1 La commande Simple Sed

Développez un programme permettant de détecter les occurrences d'un mot dans un fichier texte. et de les remplacer par un mot donné. Ce programme utilisera l'automate de localisation que vous transformerez en transducteur pour l'occasion.

```
./ssed mot remplacement fichier
```

Attention, on parle ici d'un mot au sens des langages formels. Ainsi, si l'utilisateur écrit :

```
./ssed "look like" "" shakespeare.txt
```

Le programme supprime le mot *look like* dans le fichier *shakespeare.txt*.

2 La commande ACSed

Reprendre le programme précédent et remplacer l'utilisation de l'automate de localisation par l'algorithme de Aho-Corasick. Le programme lit désormais une liste de mots dans un fichier (une ligne contient un mot à chercher et le mot par lequel le remplacer) et effectue les chercher/remplacer dans un fichier texte.

Exemple : ma my mai may mais but maison home maisons homes

En cas d'ambiguïté, le mot le plus long sera sélectionné. Vous devrez réfléchir à comment implémenter le transducteur, de façon à ce que l'espace mémoire utilisé et le temps de calcul soit les plus petits possibles. Par exemple, comment éviter de stocker deux fois le mot/préfixe "home" pour les mots *maisons* et *maisons* Cette réflexion devra apparaître dans votre rapport.

3 Rendu

Le projet est à rendre par mail à la date du Dimanche 04 décembre au soir. Le projet peut-être effectué seul ou en binôme, mais il est interdit de faire des groupes de 3 ou plus.

Le projet sera envoyé sous la forme d'un fichier *.tar.gz* contenant :

- Évitez au maimum de dupliquer du code. Réutilisez intelligemment le code d'un programme dans le programme suivant.
- Un fichier *notes.txt* dans lequel vous expliquerez comment compiler et exécuter le programme
- Le code source, les exemples et les fichiers exécutables devront être placés dans des répertoires distincts.
- Vos programmes doivent être commentés proprement.
- Un rapport de quelques pages dans lequel vous décrirez le projet, ce que qui vous a posé problème et lorsque c'est le cas, comment vous avez résolu ce problème. Le rapport ne doit pas contenir de code, mais peut faire référence à certaines fonctions dans le code.
- **La moindre ligne de code récupérée sur internet ou dans les sources d'un autre binôme sera considérée comme une tentative de fraude et l'intégralité du projet sera noté 0.** N'ayez aucun doute sur ma capacité à détecter ces fraudes.

Tout manquement à cette liste sera pénalisé et aura un impact sur la note finale.