

Programmation Shell: occurrence de mots dans un texte

Alain Aubord
HEG

27 octobre 2024

Il est souvent intéressant de savoir combien de fois un mot apparaît dans un texte. Cela peut assez facilement être réalisé avec un programme Shell. Surtout si le langage Shell dispose des tableaux associatifs (où l'indice permettant d'accéder à un élément du tableau est une chaîne de caractères). C'est le cas pour les shells Bash et KSH93 (qui sont utilisés pour le cours).

Afin de se faciliter la tâche, on effectuera les opérations suivantes sur le fichier:

1. tous les caractères seront transformés en minuscules;
2. tous les signes de ponctuation seront remplacés par des espaces;
3. tous les espaces consécutifs seront remplacés par un seul espace;
4. l'analyse des mots du texte se fera sur le fichier transformé.

Ces transformations ne devront pas être conservées dans un fichier intermédiaire mais directement utilisées dans votre script (au moyen d'une utilisation judicieuse des tuyaux et des boucles).

Pour réaliser cet exercice vous aurez probablement besoin du programme «tr» et de deux tableaux associatifs: l'un pour contenir les mots du texte, l'autre pour en contenir les occurrences.

Les indices permettant d'accéder au contenu des tableaux seront dans les deux cas les mots du texte dont on veut calculer le nombre d'occurrences. Il sera toutefois nécessaire de préfixer ces mots par une chaîne de caractères débutant par une lettre (par exemple la chaîne «Mot_») afin d'éviter les problèmes avec les mots qui ne sont constitués que de chiffres et qui seront considérés comme des nombres par le Shell.

Afin de tester votre programme, vous pourrez utiliser le document «histoire_du_soldat.txt» qui contient le texte de la pièce musicale et parlée appelée «l'histoire du soldat» écrite par Igor Stravinsky et C.F. Ramuz au début du siècle passé.

Ce texte est un beau (et assez pessimiste) conte philosophique sur la nature humaine. Le fichier pdf vous permettra de le lire confortablement si vous le souhaitez.

Les liens suivants sont des vidéos soit du spectacle soit décrivant les différents événements qui sont survenus autour de la création du spectacle.

- <https://www.rts.ch/archives/tv/culture/preface/8592001-lhistoire-du-soldat.html>
- https://www.youtube.com/watch?v=odTpm10_NI4
Une version qui semble très bien cotée.
- <https://www.youtube.com/watch?v=DnFRwSAzf30>
Seulement quelques extraits mais avec les protagonistes de la création. Probablement la plus authentique.
- Une dernière version très intéressante avec Ernest Ansermet et l'Orchestre de la Suisse Romande n'est malheureusement pas disponible en ligne.
- <https://www.youtube.com/watch?v=oZV2S5CKxjs>
Pour ceux qui veulent une version très parisienne.

Comme deuxième test, vous pourrez utiliser votre propre programme. Avant cela, faites-en une copie pour le cas où le test du programme le détruirait.