

Initiation à l'automatisation des tâches
LI218
Premier devoir à la maison

Le devoir est à rendre pour le 26 octobre 2010 avant 13h30 **impérativement** sur le site <http://www-desir.lip6.fr/li218-2010oct/soumission/soumission.php> sous la forme d'une archive tar-compressée des 4 scripts de nom *numero_etudiant.tgz*

1 Dictionnaire

On appellera ici *dictionnaire* un fichier non vide qui contient des mots formés de caractères ASCII (non accentués), triés alphabétiquement, à raison d'exactly un mot par ligne, sans aucun blanc superflu et sans ligne blanche.

Le répertoire `/Infos/lmd/2010/licence/ue/li218-2010oct/Devoir1/` contient quatre fichiers:

- le fichier `dico_egalite.txt` ne contient que trois mots.
- le fichier `microdico.txt` est un petit dictionnaire d'une centaine de mots.
- le fichier `minidoc.txt` est un petit dictionnaire d'environ 1350 mots.
- le fichier `dicofr.txt.gz` est (la forme compressée d') un dictionnaire qui contient plus de trois cent mille mots français.

Question 1 – Séparation des initiales

Écrire un script `sepi` qui prend un dictionnaire en argument et produit un fichier contenant les mêmes mots, dans le même ordre, où une ligne blanche a été insérée à chaque changement d'initiale, c'est-à-dire entre les mots commençant par la lettre a et ceux commençant par la lettre b, etc.

Par exemple:

```
% sepi microdico.txt | head -n 15
abandonner
accord
affaire
anglais
appartenir
arriere
attaquer
aussitot
avenir

bientot
bureau

casser
cesser
```

Vous écrirez deux versions du script `sepi`:

`sepi_while` qui effectue le traitement prévu en parcourant le dictionnaire avec une boucle `while`

sepi_for_grep qui effectue le traitement prévu en parcourant l'alphabet et en extrayant avec la commande `grep` les lignes commençant par chacune des initiales. Les initiales ne correspondant à aucun mot du dictionnaire ne devront pas donner lieu à une ligne blanche. La seule ligne blanche tolérée sera en fin de fichier.

Question 2 – Initiale la plus fréquente

Écrire un script `freqi` qui affiche l'initiale la plus fréquente dans le dictionnaire passé en argument, c'est-à-dire la lettre par laquelle commencent le plus de mots.

En cas d'égalité, on affichera la dernière initiale.

Par exemple:

```
% freqi microdico.txt
p
% freqi dico_egalite.txt
c
```

2 Texte

On considère à présent un texte quelconque formé de lignes pouvant contenir plusieurs mots, sans ponctuation. Vous pourrez effectuer vos essais sur les fichiers `texte` et `texte2` du répertoire `/Infos/lmd/2010/licence/ue/li218-2010oct/Devoir1/`.

Question 3 – Ligne avec le plus de mots

Écrire un script `max_mots` qui indique le numéro des lignes qui contiennent le plus grand nombre de mots.

Par exemple:

```
% max_mots texte
19
% max_mots texte2
6 19
```