



TP03 – Expr. Régulières / Shell Scripting

Systèmes d'exploitation / Classes T-2adfg

1. Téléchargez les fichiers «francais.txt.gz» et «german.txt.gz» de la forge git (<https://gitlab.forge.hefr.ch/sup-os-2016-2017/labo/tree/master/tp06>) dans le répertoire de travail de votre machine Linux (/tp/tp6).

- (a) Affichez tous les mots français de dix lettres comportant au moins cinq fois la lettre «s» et trois fois la lettre «i».

Solution :

```
$ zcat francais.txt.gz | \
  egrep '^.{10}$' | \
  egrep '(s.*){5}' | \
  egrep '(i.*){3,}'
```

- (b) Cherchez un mot français avec une même lettre qui se répète au moins 3 fois.

Solution :

```
$ zcat francais.txt.gz | egrep '(.)\1{2,}'
```

- (c) En Allemand, les mots avec la même lettre qui se répète trois fois sont courants. Combien y-en a-t-il ?

Solution :

```
$ zcat german.txt.gz | egrep '(.)\1{2,}' | wc -l
> 2823
```

- (d) Et est-ce qu'il y a des mots avec la même lettre qui se répète quatre fois ?

Solution :

```
$ zcat german.txt.gz | egrep '(.)\1{3,}'
# --> non
```

- (e) Cherchez un mot allemand de douze lettres avec une même lettre qui se répète au moins 3 fois et cette lettre ne doit pas être parmi «e, f, i, l, m, n, p, r, s t»

Solution :

```
$ zcat german.txt.gz | egrep '^.{12}$' | egrep -i '([^\efilmnprst])\1{2,}'
> Schrubbbsen
```

- (f) Affichez la liste des mots allemand commençant par un «w» et qui utilise neuf fois la même lettre.

Solution :

```
$ zcat german.txt.gz | egrep -i '^w' | egrep '(.).*\1{8}'
> Wissenschaftsausschusses
```

(g) Cherchez les mots français le plus court avec cinq fois la même lettre.

Solution :

```
$ zcat francais.txt.gz | egrep '(.)(.*\1){4}' | while read line; do \
  count=$(echo $line | wc -c); echo $count $line; \
done | sort -n | head -10
> 9 assisses
> 9 sassasse
> 10 assassins
```

2. Téléchargez aussi fichiers «collaborateurs_0.csv» et «etudiants_0.csv». Ces fichiers utilisent les caractères «latin-1» ou «ISO/IEC 8859-1». Commencez par convertir ces fichiers en unicode (UTF-8). Vous pouvez utiliser la commande «iconv» pour cette opération. Convertissez aussi les fins de lignes «\r» par «\n» avec la commande «tr». Sauvez le résultat dans les fichiers «collaborateurs.csv» et «etudiants.csv»

Solution :

```
$ iconv -f latin1 -t utf-8 collaborateurs_0.csv | tr '\r' '\n' > collaborateurs.csv
$ iconv -f latin1 -t utf-8 etudiants_0.csv | tr '\r' '\n' > etudiants.csv
```

Ces derniers fichiers sont dans le format «CSV» pour «Comma Separated Values». Dans notre exemple, les champs sont cependant séparés par des points-virgules et non des virgules. Pour extraire un champs donné, vous pouvez utiliser la commande cut. Essayez la commande suivante :

```
$ cat collaborateurs.csv | cut -d ';' -f "1,2,5"
```

(a) Parmi les collaborateurs, quel bureau héberge le plus de monde.

Solution :

```
$ cat collaborateurs.csv | cut -d ';' -f 4 | egrep '.' | sort | \
  uniq -c | sort -rn | head -n 4
> 23 B40.11
> 15 Bea/12
> 9 Fonderie 8
> 8 D30.19
```

(b) Qui se trouve dans ce bureau ?

Solution :

```
$ cat collaborateurs.csv | cut -d ';' -f "1,2,4" | egrep 'B40\..11$' | \
  cut -d ';' -f "1,2"
```

(c) Quel étudiant à un bureau ?

Solution :

```
$ cat etudiants.csv | cut -d ';' -f "1,2,4" | egrep -v ';$'
```

3. Le script suivant produit une liste de téléphone. Recopiez-le et expérimentez.

```
#!/bin/bash

input="collaborateurs.csv"
while read line; do
    phone=$(echo $line | cut -d \; -f 3)
    if [ -n "$phone" ]; then
        firstname=$(echo $line | cut -d \; -f 2)
        name=$(echo $line | cut -d \; -f 1)
        echo "${firstname} ${name} -> ${phone}"
    fi
done < ${input}
```

Modifiez le script pour faire une liste des collaborateurs qui sont seuls dans leur bureau.

Solution :

```
#!/bin/bash

input="collaborateurs.csv"
while read line; do
    office=$(echo $line | cut -d \; -f 4)
    if [ -n "$office" ]; then
        count=$(cat ${input} | cut -d \; -f 4 | egrep "^${office}\$" | wc -l)
        if [ "${count}" -eq 1 ]; then
            firstname=$(echo $line | cut -d \; -f 2)
            name=$(echo $line | cut -d \; -f 1)
            echo "${firstname} ${name} is alone in his/her office"
        fi
    fi
done < ${input}
```