

Egrep et Expressions Régulières

Introduction à UNIX

Egrep

Introduction

- Recherche dans des fichiers
 - D'un chaîne ou d'une sous chaîne de caractères
 - Simplement d'un mot
 - D'une chaîne *formalisée* par une expression régulière
- Utilisation et informations
 - **egrep** [options] <chaîne recherchée> <fichier>
 - **man egrep**
- Résultat
 - Lignes du fichier contenant ce qui est recherché
 - Ou autre résultat, suivant les options utilisées

2

Egrep

Introduction

- Il existe en réalité deux versions :
 - **grep** qui supporte simplement les expressions régulières de base
 - **egrep** qui supporte les expressions régulières dites étendues et qui offre donc plus de puissance
- La commande **egrep** supporte toutes les options de la commande **grep**
- Nous utiliserons donc dans ce cours **egrep**

3

Egrep

Recherche d'une Chaîne

- **egrep** <chaîne> fichier

Juchée à l'extrémité **nord** est de la table qui me sert de bureau, une lampe imbibe de sa lumière tamisée les quelques papiers privilégiés qui n'ont pas encore leur place dans le fatras des classeurs habitant l'étagère à laquelle je tourne le dos avec une feinte obstination.

```
biotope% egrep nord lettre.txt
Juchée à l'extrémité nord est de la table qui me sert de bureau,
```

4

Egrep

Recherche d'une Chaîne

- `egrep <chaîne> fichier`

Juchée à l'extrémité **nord**-est de la table qui me sert de bureau, une lampe imbibe de sa lumière tamisée les quelques papiers privilégiés qui n'ont pas encore **leur** place dans le fatras des classeurs habitant l'étagère à laquelle je tourne le dos avec une feinte obstination.

```
biotope% egrep or lettre.txt
```

```
Juchée à l'extrémité nord-est de la table qui me sert de bureau,
privilégiés qui n'ont pas encore leur place parmi le fatras des
```

Egrep

Recherche *Inversée*

- `egrep -v <chaîne> fichier`

Juchée à l'extrémité nord-est de la table qui me sert de bureau, une lampe imbibe de sa lumière tamisée les **quelques** papiers privilégiés qui n'ont pas encore leur place parmi le fatras des classeurs habitant l'étagère à laquelle je tourne le dos avec une feinte obstination.

```
biotope% egrep -v que lettre.txt
```

```
Juchée à l'extrémité nord-est de la table qui me sert de bureau,
privilégiés qui n'ont pas encore leur place dans le fatras des
feinte obstination.
```

Egrep

Recherche d'un Mot Exact

- `egrep -w <chaîne> fichier`

Juchée à l'extrémité nord-est de la table qui me sert de bureau, une lampe imbibe de sa lumière tamisée **les** quelques papiers privilégiés qui n'ont pas encore **leur** place parmi **le** fatras des classeurs habitant l'étagère à laquelle je tourne **le** dos avec une feinte obstination.

```
biotope% egrep -w le lettre.txt
```

```
privilégiés qui n'ont pas encore leur place dans le fatras des
classeurs habitant l'étagère à laquelle je tourne le dos avec une
```

Egrep

Lignes de Contexte

- `egrep -<nombre de lignes> <chaîne> fichier`

Juchée à l'extrémité nord-est de la table qui me sert de bureau,
 → une lampe imbibe de sa lumière tamisée les quelques papiers
 privilégiés qui n'ont pas encore **leur** place dans le fatras des
 → classeurs habitant l'étagère à laquelle je tourne le dos avec une
 feinte obstination.

```
biotope% egrep -l leur lettre.txt
```

```
→ une lampe imbibe de sa lumière tamisée les quelques papiers
privilégiés qui n'ont pas encore leur place parmi le fatras des
→ classeurs habitant l'étagère à laquelle je tourne le dos avec une
```

Egrep

Numéro de Ligne

- `egrep -n <chaîne> fichier`

Juchée à l'extrémité nord-est de la table qui me sert de bureau, une lampe imbibe de sa lumière tamisée les quelques papiers privilégiés qui n'ont pas encore leur place dans le fatras des classeurs habitant l'étagère à laquelle je tourne le dos avec une feinte obstination.

```
biotope% egrep -n leur lettre.txt
```

```
3:privilégiés qui n'ont pas encore leur place parmi le fatras des
```

Egrep

Combinaison d'Options

- `egrep -n<nombre de lignes> <chaîne> fichier`

Juchée à l'extrémité nord-est de la table qui me sert de bureau, → une lampe imbibe de sa lumière tamisée les quelques papiers privilégiés qui n'ont pas encore leur place parmi le fatras des → classeurs habitant l'étagère à laquelle je tourne le dos avec une feinte obstination.

```
biotope% egrep -n1 leur lettre.txt
```

```
→ 2-une lampe imbibe de sa lumière tamisée les quelques papiers
```

```
3:privilégiés qui n'ont pas encore leur place parmi le fatras des
```

```
→ 4-classeurs habitant l'étagère à laquelle je tourne le dos avec une
```

Egrep

Respect de la Casse

- `egrep -i <chaîne> fichier`

Juchée à l'extrémité nord-est de la table qui me sert de bureau, une lampe imbibe de sa lumière tamisée les quelques papiers privilégiés qui n'ont pas encore leur place parmi le fatras des classeurs habitant l'étagère à laquelle je tourne le dos avec une feinte obstination.

```
biotope% egrep j lettre.txt
```

```
classeurs habitant l'étagère à laquelle je tourne le dos avec une
```

```
biotope% egrep -i j lettre.txt
```

```
Juchée à l'extrémité nord-est de la table qui me sert de bureau, classeurs habitant l'étagère à laquelle je tourne le dos avec une
```

Egrep

Compte des occurrences

- `egrep -c <chaîne> fichier`

Juchée à l'extrémité nord-est de la table qui me sert de bureau, une lampe imbibe de sa lumière tamisée les quelques papiers privilégiés qui n'ont pas encore leur place parmi le fatras des classeurs habitant l'étagère à laquelle je tourne le dos avec une feinte obstination.

```
biotope% egrep -c que lettre.txt
```

```
2
```

Egrep

Exemple Concret

- Imaginons le fichier de notes suivant

```
crepetna:Crepet, Nathalie:CREN1807750:92:87:88:54:70
yosnheat:Yos Nhean, Trakal:YOST19087603:84:73:70:50:73
benelaur:Benel, Aurelien:BENA80207700:84:73:89:45:100
soucypas:Soucy, Pascal:SOUP14067502:95:90:89:87:99
```

- On peut extraire les lignes qui contiennent une note comprise entre 90 et 99

```
biotope% egrep :9 notes.txt
crepetna:Crepet, Nathalie:CREN1807750:92:87:88:54:70
soucypas:Soucy, Pascal:SOUP14067502:95:90:89:87:99
```

Egrep

Exemple Concret

```
biotope% egrep :9 notes.txt
crepetna:Crepet, Nathalie:CREN1807750:92:87:88:54:70
soucypas:Soucy, Pascal:SOUP14067502:95:90:89:87:99
```

- En combinant la commande avec **cut**, on peut extraire les noms des personnes concernées

```
biotope% egrep :9 notes.txt | cut -d: -f2
Crepet, Nathalie
Soucy, Pascal
```

Egrep

Limite ?

```
crepetna:Crepet, Nathalie:CREN1807750:92:87:88:54:70
yosnheat:Yos Nhean, Trakal:YOST19087603:84:73:70:50:73
benelaur:Benel, Aurelien:BENA80207700:84:73:89:45:100
soucypas:Soucy, Pascal:SOUP14067502:95:90:89:87:99
```

- Comment faire pour extraire les lignes où la dernière note est comprise entre 90 et 99 ?
- Solution : Utiliser une expression régulière
- Mais, qu'est ce qu'une expression régulière ?

Expressions Régulières Introduction

- Définition
 - Formule qui représente une chaîne de caractères
 - Composée de caractères et d'opérateurs
- Utilisation
 - On recherche alors non pas un mot ou une simple chaîne de caractères mais une suite de caractères qui correspondent aux critères énoncés par la formule
 - Certains opérateurs doivent être précédés d'un \ pour ne pas entrer en conflit avec le shell, ainsi : {, }, <, >, (,) et | seront écrits \{, \}, \<, \>, \(, \) et \| . C'est aussi le cas de l'espace
- Culture Générale
 - Expression régulière, "regular expression" en anglais, se traduit en bon français par "expression rationnelle" mais l'usage est de dire "régulière"

Expressions Régulières Opérateurs

- Comment représenter
 - Un caractère quelconque : `.`
 - Une ou une infinité d'occurrences : `+`
 - Zéro ou une infinité d'occurrences : `*`
 - Un choix parmi un ensemble : `[<liste>]`
 - Tout sauf un certain caractère : `[^<caractère>]`
- Evidement, on peut combiner les expressions
 - `.*` : Zéro ou une infinité de caractères quelconques
 - `a+b*` : Au moins un 'a' suivi de 0 ou une infinité de 'b'
 - `[ab]+` : Au moins un 'a' ou 'b' ou une infinité
 - etc.

17

Expressions Régulières Utilisation

- Exemples
 - `v.+` : Les chaînes contenant un v suivi de n'importe quelle suite de caractères
 - vandale
 - vestale
 - lavage
 - `[vs].+` : Les chaînes contenant un 'v' ou un 's' suivi de n'importe quelle suite de caractères
 - vandale
 - voiture
 - descendre
 - sandale

18

Expressions Régulières Utilisation

- Exemples
 - `a.*a` : Les chaînes contenant deux 'a'
 - palabres
 - sandale
 - pascale
 - cascade
 - `[ps].*a.*a` : Les chaînes contenant un 'p' ou un 's' suivi d'une sous chaîne contenant deux 'a'
 - sandale
 - pascale
 - apprentissage linéaire

19

Egrep Exemple

- Reprenons le fichier de notes précédent


```
crepetna:Crepet, Nathalie:CREN1807750:92:87:88:54:70
yosnheat:Yos Nhean, Trakal:YOST19087603:84:73:70:50:73
benelaure:Benel, Aurelien:BENA80207700:84:73:89:45:100
soucypas:Soucy, Pascal:SOUP14067502:95:90:89:87:99
```
- On peut maintenant extraire la ligne où la dernière note est comprise entre 90 et 99
 - Une chaîne avec 7 ':' entre lesquels on peut avoir n'importe quoi et dont le dernier ':' est suivi d'un 9

```
biotope% egrep '.*:.*:.*:.*:.*:.*:.*:9' notes.txt
soucypas:Soucy, Pascal:SOUP14067502:95:90:89:87:99
```

20

Expressions Régulières Opérateurs

- Comment représenter
 - Un caractère compris entre 'a' et 'z' : **[a-z]**
 - Un caractère compris entre '0' et '9' : **[0-9]**
 - Le début d'une ligne : **^**
 - La fin d'une ligne : **\$**
 - Un choix entre deux chaînes : **|**
 - Et en combinant avec le choix
 - Caractère compris entre 'a' et 'z' ou 'A' et 'Z' : **[a-zA-Z]**
- Répétitions d'une occurrence
 - Exactement **2** répétitions de 'x' : **x{2}**
 - Entre **2** et **5** répétitions de 'x' : **x{2,5}**
 - Au moins **2** répétitions de 'x' : **x{2,}**

21

Egrep Exemple

- Encore le fameux fichier de notes

```
crepetna:Crepet, Nathalie:CREN1807750:92:87:88:54:70
yosnheat:Yos Nhean, Trakal:YOST19087603:84:73:70:50:73
benelaur:Benel, Aurelien:BENA80207700:84:73:89:45:100
soucypas:Soucy, Pascal:SOUP14067502:95:90:89:87:99
```

- Aviez vous remarqué que le code permanent de Nathalie ne contient que 7 chiffres au lieu des 8 habituels ?
- On peut facilement identifier cela en utilisant une expression régulière

22

Egrep Exemple

- Il suffit de *formaliser* le code permanent

```
crepetna:Crepet, Nathalie:CREN1807750:92:87:88:54:70
yosnheat:Yos Nhean, Trakal:YOST19087603:84:73:70:50:73
benelaur:Benel, Aurelien:BENA80207700:84:73:89:45:100
soucypas:Soucy, Pascal:SOUP14067502:95:90:89:87:99
```

- Un code permanent contient **4 majuscules** suivies de **8 chiffres** : **[A-Z]{4}[0-9]{8}**

```
biotope% egrep [A-Z]{4}[0-9]{8} notes.txt
yosnheat:Yos Nhean, Trakal:YOST19087603:84:73:70:50:73
benelaur:Benel, Aurelien:BENA80207700:84:73:89:45:100
soucypas:Soucy, Pascal:SOUP14067502:95:90:89:87:99
```

23

Egrep Exemple

- On peut aussi ne faire apparaître que la ligne qui est fautive

```
crepetna:Crepet, Nathalie:CREN1807750:92:87:88:54:70
yosnheat:Yos Nhean, Trakal:YOST19087603:84:73:70:50:73
benelaur:Benel, Aurelien:BENA80207700:84:73:89:45:100
soucypas:Soucy, Pascal:SOUP14067502:95:90:89:87:99
```

- Simplement en utilisant l'option **-v**

```
biotope% egrep -v [A-Z]{4}[0-9]{8} notes.txt
crepetna:Crepet, Nathalie:CREN1807750:92:87:88:54:70
```

24

Egrep

Exemple

- Autre solution possible

```
crepetna:Crepet, Nathalie:CREN1807750:92:87:88:54:70
yosnheat:Yos Nhean, Trakal:YOST19087603:84:73:70:50:73
benelaur:Benel, Aurelien:BENA80207700:84:73:89:45:100
soucypas:Soucy, Pascal:SOUP14067502:95:90:89:87:99
```

- Finalement, un code permanent contient 12 caractères lettres ou chiffres : `[A-Z0-9]{12}`

```
biotope% egrep [A-Z0-9]{12} notes.txt
yosnheat:Yos Nhean, Trakal:YOST19087603:84:73:70:50:73
benelaur:Benel, Aurelien:BENA80207700:84:73:89:45:100
soucypas:Soucy, Pascal:SOUP14067502:95:90:89:87:99
```

Egrep

Exemple

- Autre solution pour la dernière note

```
crepetna:Crepet, Nathalie:CREN1807750:92:87:88:54:70
yosnheat:Yos Nhean, Trakal:YOST19087603:84:73:70:50:73
benelaur:Benel, Aurelien:BENA80207700:84:73:89:45:100
soucypas:Soucy, Pascal:SOUP14067502:95:90:89:87:99
```

- Une chaîne composée de 6 fois ':' suivi d'une chaîne quelconque puis d'un ':9'. Donc : `\(:.*\){6}:9`

```
biotope% egrep \(:.*\){6}:9 notes.txt
soucypas:Soucy, Pascal:SOUP14067502:95:90:89:87:99
```

Egrep

Exemple

- Lignes commençant par un 'p' ou un 'c'

Juchée à l'extrémité nord-est de la table qui me sert de bureau, une lampe imbibe de sa lumière tamisée les quelques papiers privilégiés qui n'ont pas encore leur place parmi le fatras des classeurs habitant l'étagère à laquelle je tourne le dos avec une feinte obstination.

```
biotope% egrep ^[pc] lettre.txt
privilégiés qui n'ont pas encore leur place dans le fatras des
classeurs habitant l'étagère à laquelle je tourne le dos avec une
```

Egrep

Exemple

- Lignes finissant par un 's'

Juchée à l'extrémité nord-est de la table qui me sert de bureau, une lampe imbibe de sa lumière tamisée les quelques papiers privilégiés qui n'ont pas encore leur place parmi le fatras des classeurs habitant l'étagère à laquelle je tourne le dos avec une feinte obstination.

```
biotope% egrep s$ lettre.txt
une lampe imbibe de sa lumiere tamisee les quelques papiers
privilégiés qui n'ont pas encore leur place dans le fatras des
```

Egrep

Exemple

- Recherche des lignes contenant 'leur' ou 'qui'

Juchée à l'extrémité nord-est de la table qui me sert de bureau, une lampe imbibe de sa lumière tamisée les quelques papiers privilégiés qui n'ont pas encore leur place parmi le fatras des classeurs habitant l'étagère à laquelle je tourne le dos avec une feinte obstination.

```
biotope% egrep \(leur\|qui\) lettre.txt
Juchée à l'extrémité nord-est de la table qui me sert de bureau,
privilégiés qui n'ont pas encore leur place parmi le fatras des
```

Egrep

Exemple Complexe

- Repérer les lignes contenant une date qui est exprimée par JJ/MM/AA ou bien JJ/MM/AAAA
- Formalisation des jours
 - Les jours sont compris entre 1 et 31
 - Le premier chiffre est compris entre 0 et 3
 - Le deuxième chiffre est compris entre 0 et 9 sauf dans les cas où le premier est un 0 ou bien un 3
 - On obtient alors : $\backslash(0[1-9]\backslash[1-2][0-9]\backslash3[0-1])\backslash$
 - Jours 01 à 09 : $0[1-9]$
 - Jours 10 à 29 : $[1-2][0-9]$
 - Jours 30 et 31 : $3[0-1]$

Egrep

Exemple Complexe

- Formalisation des mois
 - Similairement, les mois sont compris entre 1 et 12
 - Le premier chiffre est compris entre 0 et 1
 - Le deuxième chiffre est compris entre 0 et 9 sauf dans les cas où le premier est un 0
 - On obtient alors : $\backslash(0[1-9]\backslash1[0-2])\backslash$
 - Mois 01 à 09 : $0[1-9]$
 - Mois 10 à 12 : $1[0-2]$
- Formalisation des années
 - Une année contient simplement 2 ou 4 chiffres
 - On obtient alors : $\backslash([0-9]\backslash2\backslash)\backslash([0-9]\backslash4\backslash)\backslash$

Egrep

Exemple Complexe

- Formalisation de la date
 - Concaténation des expressions formalisant les jours, mois et années avec un '/' entre elles :

$$\backslash(0[1-9]\backslash[1-2][0-9]\backslash3[0-1])\backslash(0[1-9]\backslash1[0-2])\backslash([0-9]\backslash2\backslash)\backslash([0-9]\backslash4\backslash)\backslash$$
- Problème ?
 - Et oui, nous n'avons pas tenu compte du fait que certains mois ont 30 jours et d'autres 31.
 - Nous n'avons pas tenu compte du mois de février qui ne peut avoir que 28 ou 29 jours, ni du fait qu'il a 29 jours que lors des années bissextiles.
- La solution vous est laissée en exercice ;-)