

3.4 - Trier et transformer avec sort, uniq, cut, sed et awk

Les petits outils Unix sont conçus pour être combinés. Ils permettent d'extraire une information sans ouvrir un tableur ou écrire un programme complet.

Compter avec wc

```
wc fichier.txt
wc -l fichier.txt      # lignes
wc -w fichier.txt      # mots
wc -c fichier.txt      # octets
```

Trier et supprimer les doublons

```
sort noms.txt
sort -n nombres.txt
sort -h tailles.txt
sort -r noms.txt
sort noms.txt | uniq
sort noms.txt | uniq -c | sort -nr
```

`uniq` ne regroupe que les lignes identiques adjacentes ; il est donc souvent précédé de `sort`.

Extraire des colonnes

```
cut -d: -f1 /etc/passwd  
cut -d';' -f1,3 inventaire.csv
```

`cut` convient aux séparateurs simples. Un CSV avec des guillemets, des retours à la ligne ou des séparateurs échappés nécessite un véritable analyseur CSV.

Remplacer des caractères avec tr

```
printf '%s\n' 'bonjour' | tr '[:lower:]' '[:upper:]'  
tr -d '\r' < windows.txt > linux.txt  
tr ':' '\n' < /etc/passwd
```

Transformer avec sed

Afficher sans modifier :

```
sed -n '1,10p' fichier.txt  
sed 's/ancien/nouveau/g' fichier.txt  
sed '/^#/d; /^$/d' fichier.conf
```

Créer un nouveau fichier :

```
sed 's/ancien/nouveau/g' fichier.txt > fichier-modifie.txt
```

La modification directe `sed -i` varie entre GNU/Linux et BSD/macOS. Créer d'abord une sauvegarde :

```
sed -i.bak 's/ancien/nouveau/g' fichier.txt
```

Extraire avec awk

```
awk -F: '{print $1}' /etc/passwd  
awk '{print NR, $0}' fichier.txt  
awk '$3 > 80 {print $1, $3}' mesures.txt  
df -P | awk 'NR>1 {print $6, $5}'
```

Dans `awk`, `$1` désigne le premier champ, `$0` toute la ligne et `NR` le numéro de ligne.

Exemple combiné

Compter les shells déclarés dans `/etc/passwd` :

```
cut -d: -f7 /etc/passwd | sort | uniq -c | sort -nr
```

Tester d'abord les transformations sans `-i`, comparer le résultat avec `diff`, puis remplacer le fichier uniquement après validation.

Revision #2

Created 2026-08-08 01:24:50 CEST by Admin

Updated 2026-08-08 01:46:05 CEST by Admin