

Introduction à AWK

Outil et langage
de manipulation de texte

Histoire

- Créé dans les années 1970 par
 - Alfred Aho
 - Peter Weinberger
 - Brian Kernighan
- AWK est un langage
- awk est le nom de l'outil qui exécute AWK
- Hormis le shell, le seul langage scripté présent dans un environnement standard UNIX
- AWK est un des points de départ de Perl

Finalités

- Parseur de fichiers structurés (XML, CSV...)
- Transformations et mises en forme avancées
- Re-génération de texte (requêtes SQL)
- Extraction d'informations
- Calculs et statistiques sur le texte

Fonctionnement

- Lit chaque ligne une à une
- Découpe le fichiers en 2 dimensions :
 - RECORD (ligne par défaut)
 - FIELD (mot par défaut, séparé par un espace)

\$0	\$1	\$2	\$3	\$4	\$5	\$6	\$7	\$8	\$9	\$10	\$11	\$12	\$13
NR 1	Plus	il	y	a	de	fromage,	plus	il	y	a	de	trous;	
NR 2	or	plus	il	y	a	de	trous,	moins	il	y	a	de	fromage;
NR 3	donc	plus	il	y	a	de	fromage,	moins	il	y	a	de	fromage.

- NR est le compteur de record, NF le compteur de champ

Structure

- Chaque record est testé contre chaque condition
- Une action est exécutée si la condition est vérifiée
`awk 'condition { action } condition { action } ... ' file`
- Par défaut : condition true, action print
`{ print $1 } # affiche chaque premier mot de chaque ligne`

```
$0 ~ /Important/ { print }  
length($0) > 80 { print "LIGNE " NR " TROP LONGUE" }  
# affiche chaque ligne qui contient "Important"  
# et "LIGNE n TROP LONGUE" si n-ième ligne > 80 caractères
```

Variables et éléments de langage

- BEGIN, END: exécute ces blocs avant et après le reste
- NR : compte du nombre d'enregistrements (lignes)
- NF : compte du nombre de champs (mots)
- RS : record separator changé avec BEGIN { RS=":" }
- FS : field separator, changé directement avec awk -F

```
{ # affiche à la fin le nombre total de lignes, mots et caractères :  
    w += NF  
    c += length  
}  
END { print NR, w, c }
```

Exemple avec if / else

Mettre la dernière date rencontrée au début de chaque ligne

1955

The Musings of Miles

Blue Moods

Miles Davis Quintet

1956

Cookin' with the Miles Davis Quintet

1957

Relaxin' with the Miles Davis Quintet

Steamin' with the Miles Davis Quintet

```
{
  if ( $0 ~ /^[0-9]+$/ ) {
    year = $0
  } else {
    print year " : " $0
  }
}
```

Exemple avec getline

Dupliquer les blocs de traductions en_US pour en_GB

Country:

C1:

iso_code: AFG

Translation:

en_US:

name: Afghanistan

slug: afghanistan

C2:

iso_code: ZAF

Translation:

en_US:

name: South Africa

name: south-africa

```
{
  if ($0 ~ /en_US/) {
    print $0
    line = "" # (ré)initialise la variable line
    while (getline && $0 ~ /[a-z]/) { # tant qu'on a pas de ligne vide
      line = line "\n" $0 # ajoute la ligne à la variable
      print $0 # on affiche de toute façon la ligne pour le bloc en_US
    };
    print "      en_GB:" line "\n" # on affiche le contenu pour en_GB
  } else {
    print $0 # autre type de ligne, on affiche
  }
}
```

Résultat

Country:

C1:

iso_code: AFG

Translation:

en_US:

name: Afghanistan

slug: afghanistan

en_GB:

name: Afghanistan

slug: afghanistan

C2:

iso_code: ZAF

Translation:

en_US:

name: South Africa

slug: south-africa

en_GB:

name: South Africa

slug: south-africa