

# awk et les expressions régulières

- Objet : utiliser les expressions régulières avec awk
- Niveau requis : [avisé](#)
- Commentaires : *Contexte d'utilisation du sujet du tuto.*

## Détail sur les expressions régulières awk

### Les modèles de recherche sur les chaînes

Types	Significations
/ER/	Sélectionne la ligne courante lorsqu'une sous chaîne est valide en fonction de l'expression régulière
expression~/ER/	Sélectionne la ligne courante si la chaîne "expression" contient une sous chaîne valide en fonction de l'expression régulière
expression!~/ER/	Sélectionne la ligne courante si la chaîne "expression" ne contient pas une sous chaîne valide en fonction de l'expression régulière.



Si la chaîne à laquelle on désire soumettre l'expression régulière n'est pas dans la variable prédéfinie `$_` l'opérateur `=~` va nous permettre de soumettre l'expression régulière qui se trouve à sa droite à la variable qui se trouve à sa gauche. Le résultat de cette opération est "vrai" si une concordance est trouvée, "faux" sinon.

### Caractères des expressions régulières

#### Comme avec grep, sed

1) Une expression régulière est toujours placée entre deux slash.

**/expression régulière/**

2) Métacaractères: Ce sont des caractères réservés à valeur représentative :

`\ ^ $ [ ] | ( ) * + ? ^ $ . \< \>`



3) Si l'on veut exprimer un de ces caractères, il est indispensable de le faire précéder d'un antislash \.

4) Utilisation des classes :

- **[AEIOU]** : Sélection des caractères A, E, I, O, U.
- **[^AEIOU]** : Sélection de tous les caractères sauf A, E, I, O, U.
- **^[^AEIOU]** : Sélection de tous les caractères en début de chaîne sauf A, E, I, O, U.
- **^[^a-z-]** : Sélection de toute chaîne de un caractère n'étant ni une minuscule

ni un -

- **[a-zA-Z][0-9]** : Sélection de toute chaîne de deux caractères, une lettre majuscule ou minuscule suivie d'un chiffre.



A l'intérieur d'une classe de caractères, tout caractère admet sa signification propre, sauf \ et ^ placé au début, et - entre deux caractères.

- **[.]** : Sélection du caractère .
- **^[^^]** : Sélection de tout caractère en début de chaîne sauf ^.

Il y a deux opérateurs binaires sur les expressions régulières :

1. l'intersection (et)
2. l'union (ou)



Si  $r_1$  et  $r_2$  sont deux expressions régulières, alors :

- **$r_1 | r_2$**  sélectionne toute chaîne validée par  **$r_1$  ou par  $r_2$**
- **$(r_1)(r_2)$**  sélectionne toute chaîne de la forme AB, où  **$r_1$  valide A et  $r_2$  valide B**

## Les opérateurs étendus

On les appelle aussi "Les opérateurs unaires".

Les significations peuvent différer d'avec sed.

Les opérateurs + \* et ? sont utilisés pour spécifier des répétitions dans les expressions régulières.



- **$(r)^*$**  : Toute chaîne consistant en zéro ou n sous chaînes consécutives de modèle r.
- **$(r)^+$**  : Toute chaîne consistant en une ou n sous chaînes consécutives de modèle r.
- **$(r)?$**  : Toute chaîne de modèle r ou la chaîne nulle
- **$A^*$**  : La chaîne nulle ou A ou AA ou AAA ...
- **$Ab^*A$**  : La chaîne AA ou AbA ou AbbA ...
- **$Ab?A$**  : La chaîne AA ou AbA
- **$Ab+A$**  : La chaîne AbA ou AbbA ...
- **$Abb^*A$**  : La chaîne AbA ou AbbA ...
- **$[A-Z]^+$**  : Toute chaîne de une ou plusieurs lettres majuscules.
- **$(AA)^+e$**  : La chaîne AAe ou AAAe ou AAAAAe

# Application

From:

<http://debian-facile.org/> - Documentation - Wiki

Permanent link:

<http://debian-facile.org/utilisateurs:hyathie:tutos:awk-et-les-expressions-rationnelles>

Last update: **01/08/2014 19:39**

