ITESCIA
L'ECOLE DU i-MANAGEMENT

Les Expressions Régulières

❖ C’est quoi une expression régulière?

Une expression régulière est une suite de caractères typographiques (qu’on appelle plus simplement « motif » – « pattern » en anglais) décrivant un ensemble de chaînes de caractères.

→Par exemple l’ensemble de mots « ex-équo, ex-equo, ex-aequo et ex-æquo » peut être condensé en un seul motif:

« ex-(a?e|æ|é)quo ».

Les mécanismes de base pour former de telles expressions sont basés sur des caractères spéciaux de substitution, de groupement et de quantification. [Définition Wikepedia]

Expressions régulières

- Analyser des chaînes de caractères
- Pattern matching
- **Utilisation** :
  - Commandes : grep, sed, …
  - Éditeurs de textes : vi, nedit, …
  - Langages de programmation : php, perl, …

Les symboles ^ et $

^ : début d'un pattern

$: fin d'un pattern

Exemples :

^at : chaîne de caractères qui commence par « at »

cot$: chaîne de caractères qui se finit « cot »

^mot$: chaîne de caractères « mot »

test : chaîne de caractères qui contient « test »

Les symboles *, + et ?

Nombre de fois qu'un caractère (suite de caractères) puisse apparaître

- * : aucune fois ou plusieurs fois
- + : une fois ou plusieurs fois
- ? : aucune fois ou une et une seule fois

Exemples :

- ab* : chaîne de caractère contenant un a suivi d'un, de plusieurs, ou d'aucun b ("a", "ab", "abb" , …)
- ab+ : chaîne de caractère contenant un a suivi d'au moins un b ("ab", "abb", "abbb", …)
- ab? : chaîne de caractère contenant un a suivi d'un ou d'aucun b ("a", "ab ", mais pas "abb ")
- a?b+$: chaîne de caractères composée d'aucun ou d'un seul a, suivi d'un ou de plusieurs b, le tout étant situé à la fin de la chaîne

Les accolades {}

Nombre d'occurrences de la chaîne

Exemples :

- ab{2} : chaîne de caractère composée d'un a suivi d'exactement deux b ("abb")
- ab{2,} : chaîne de caractère composée d'un a suivi d'au moins deux b ("abb", "abbb",…)
- ab{,5} : chaîne de caractère composée d'un a suivi de jusqu'à cinq b ("a", "ab" … et "abbbbb")
- ab{3,5} : chaîne de caractère composée d'un a suivi de trois à cinq b ("abbb", "abbbb" et "abbbbb")

Les parenthèses ()

Quantifier une chaîne de caractères

Exemples :

- a(bc)* : chaîne de caractères commençant par un a suivi d'aucune ou de plusieurs séquence de caractères "bc"
- a(bc){1,5} : chaîne de caractères commençant par un a suivi d'une à cinq fois la séquence de caractères "bc"

Le symbole |

Comme opérateur booléen OU

Exemples :

- toto|titi : chaîne de caractères contenant le mot "toto" ou le mot "titi"
- (b|cd)ef : chaîne de caractères qui contient la séquence de caractères "bef" ou bien la séquence de caractères "cdef«

A noter: (b|cd)ef équivaut (bef|cdef)

- (a|b)*c : chaîne de caractères qui contient une alternance de a et de b, se terminant par un c ("babababbbaac", "c", "bc")

Le symbole .

N'importe quel caractère unique.

Exemple :

- ^.{3}$: chaîne de caractères comportant exactement trois caractères

Les crochets []

Les caractères permis à un endroit précis d'un modèle

Exemples :

- [ab] : chaîne de caractères contenant un "a" ou un "b"
- [a-d] : chaîne de caractères qui contient les lettres minuscules comprises entre le "a" et le "d"
- ^[a-zA-Z] : chaîne de caractères qui commence par une lettre minuscules ou bien par une lettre majuscule
- [0-9]% : chaîne de caractères qui contient un pourcentage à un seul chiffre
- ,[a-zA-Z0-9]$: chaîne de caractères qui finit par une virgule suivi d'un caractère alphanumérique

Le ^ dans les crochets

- ^ comme premier symbole dans les crochets: « tout sauf »
- ^[^a]: chaîne de caractères qui ne commence pas par « a »

Notez

Si l'on veut qu'un méta-caractère apparaisse tel quel, il faut le précéder d'un backslash

Pour que les méta-caractères ?, +, {, }, |, (, et) gagnent leurs significations spéciales dans un terminal, il faut utiliser un backslash :

- \?, \+, \{, \}, \|, \(, et \).

Exemple:

grep "^a\(b\|c\)\{1,3\}" fichier permet de reconnaitre a(b|c){1,3}

Par exemple, **[a-z]\[[0-9]\]** permet de trouver **c[8]**.

ITESCIA
L'ECOLE DU I-MANAGEMENT

Bibliographie:

- [1] Cours de **Vincent Granet (**Polytech'Nice-Sophia)
- [2] Introduction à Unix et GNU / Linux, par Michael Opdenacker (Free Electrons), Traduction française par Julien Boibessot. Mise à jour Fabien Deleu (Département GTR de l'IUT de Béthune)
- Introduction à LINUX, de M. Abdallah ELKHYARI, Univ. Jean Monet St Etie
- https://moodle.polymtl.ca/mod/url/view.php?id=47398
- Cours "Systèmes d'exploitation", Audrey Queudet, Univ Nantes, 2010

Pour aller plus loin:

Livres

- *Linux pour les nuls, Dee-Ann Leblanc*, First Interactive, 2006.
- *Linux en pratique,* Arnold Robbins, Campus Press, 2007.
- *Linux programmation système et réseau , cours exemples et exercices corrigés en C-C++*, Joëlle Delacroix, Dunod, 2007

Sites web

- http://www.linux.org/
- http://www.linux-france.org/