

Annexe A – Expressions régulières

TP disponible sur la page web personnelle d'Alexandre Sedoglavic : www.lifl.fr/~sedoglav/SHELL/

Deux spécification des méta-caractères sont couramment utilisées :

1. les **méta-caractères du shell** pour identifier des noms de fichiers. Ils sont utilisés directement par le shell et aussi par certaines commandes comme find.
2. les **méta-caractères plus généraux** utilisés par la plupart des commandes identifiant des expressions régulières.

Voir le cours pour les méta-caractères du shell.

En plus des méta-caractères vue en cours, on a

- le caractère \ transforme un méta-caractère en littéral ;
- le caractère . correspond à n'importe quel caractère ;
- le caractère ^ marque le début d'une chaîne ;
- le caractère \$ marque la fin d'une chaîne ;
- le caractère + correspond à une ou plusieurs occurrences du caractère précédent.

Ainsi, `^#!/bin/sh$` correspond uniquement à la chaîne `#!/bin/sh`, `$` reconnaît une chaîne vide, et

De plus, le caractère | marque l'alternative. Ainsi, `L[y|i]s` correspond à `Lys` ou `Lis`, `^(De|Sujet|Date):@` correspond à tout ce qui commence par `De:@` ou `Sujet:@` ou `Date:@`.

L'expression régulière

- `{x}` avec `x` un nombre correspond à `x` occurrences du caractère précédent ;
- `{x,}` correspond à `x` occurrences et plus du caractère précédent ;
- `{x,y}` correspond à au minimum `x` et au maximum `y` occurrences du caractère précédent.

Ainsi,

- `a?` correspond à 0 ou 1 `a` ;
- `a*` correspond à 0, 1 ou plusieurs `a` ;
- `a+` correspond à 1 ou plusieurs `a` ;
- `a{2}` correspond à `aa` ;
- `a{2,3}` correspond à `aa` ou `aaa`.

Les crochets `[]` indique une classe où un intervalle dans une classe.

Par exemple :

- `[a-z]` correspond à les lettres de `a` à `z` ;
- `[Bb]onjour` correspond à `Bonjour` ou `bonjour` ;
- `<h[1-6]>` une balise HTML de titre par exemple.

Le tiret dans [] représente un intervalle, si on veut l'utiliser comme littéral normal, il faut le mettre après le [ou avant le], [a-z-].

L'expression régulière [^...] au lieu de [...] indique une classe complémentée.

Par exemple,

- [^0-9] correspond à tout ce qui n'est pas des chiffres ;
- [^1-6] correspond à tout sauf les chiffres de 1 à 6.

Remarquez que lorsque ^ est entre [], sa sémantique est différente et ne signifie pas début de ligne.

L'expression régulière () encadre des instructions pour capturer des sous-chaînes.

Par exemple, (?ho)+ correspond à ho, ho ho, ho ho ho, hohoho, etc.