

Regex

Métacaractères

- `^` : Début de chaîne
- `$` : Fin de chaîne ou fin de ligne
- `.` : N'importe quel caractère
- `|` : Ou (a|b pour a ou b)
- `()` : Délimiteur de groupe
- `[]` : Délimiteur d'ensemble
- `\` : caractère d'échappement

Opérateurs de répétition

- `*` : 0 ou plus
- `+` : 1 ou plus
- `?` : 0 ou 1 fois
- `{n}` : exactement n fois
- `{n,}` : n fois ou plus
- `{n,m}` : de n à m fois

Groupe et ensembles

- `(...)` : définit un groupe : c'est une sous-expression régulière
- `[...]` : définit un ensemble : c'est une liste de caractères à rechercher
- `[^...]` : définit un ensemble négatif : c'est une liste de caractères à exclure
- `[a-z]` : définit un ensemble de caractères : c'est une plage de caractères à rechercher depuis la table ASCII
- `\n` ou `$n` : énième groupe

Ensembles prédéfinis

- `\s` : n'importe quel caractère blanc : espace, tabulation, retour à la ligne, [

- `\S` : n'importe quel caractère **non** blanc, `[^]+`
- `\d` : n'importe quel nombre, `[0-9]`
- `\D` : n'importe quel **non** nombre, `[^0-9]`
- `\w` : n'importe quel caractère alphanumérique, `[a-zA-Z0-9_]`
- `\W` : n'importe quel caractère non alphanumérique, `[^a-zA-Z0-9_]`

Caractères spéciaux

- `\n` : Nouvelle ligne
- `\r` : Retour à la ligne
- `\t` : Tabulation
- `\x{0281}` : Car. Unicode U+0281 (K)

Ancres

- `\A` : Début de chaîne, équivaut à `^`
- `\Z` : Fin de chaîne, équivaut à `$`
- `\b` : Borne de mot, marque la fin, ou le début, d'une séquence de type `\w`
- `\B` : Non borne de mot, marque tout ce que `\b` ne marque pas...

Lookahead et Lookbehind

- `(?=...)` : ce qu'il y a après (positive lookahead)
- `(?!...)` : ce qu'il y **n'a pas** après (negative lookahead)
- `(?<=...)` : ce qu'il y a avant (positive lookbehind)
- `(?<!=...)` : ce qu'il y **n'a pas** avant (negative lookbehind)