



Haute école d'ingénierie et d'architecture Fribourg
Hochschule für Technik und Architektur Freiburg

Systèmes d'exploitation

Les expressions régulières



«Une expression rationnelle, ou expression régulière (traduction de l'anglais regular expression), est, en informatique, une chaîne de caractères que l'on appelle parfois un motif et qui décrit un ensemble de chaînes de caractères possibles selon une syntaxe précise.»

Les expressions régulières



C'est un peu comme le jeu des plots à encastrer





- Les expressions régulières sont souvent notées entre deux «/»
- Une expression régulière qui ne contient que des lettres fait correspondre un texte composé des mêmes lettres.

Expression régulière	Texte accepté
/hello/	hello



- La barre verticale « | » définit un choix ou une alternative.

Expression régulière	Texte accepté
/gray grey/	gray grey



- Les parenthèses « () » permettent de définir un groupe avec une précedence plus élevée.

Expression régulière	Texte accepté
/gr(a e)y/	gray grey



- Le point d'interrogation «?» est un quantifieur qui définit que ce qui précède le point d'interrogation est optionnel. En d'autres mots, il peut exister zéro ou une fois.

Expression régulière	Texte accepté
/colou?r/	colour color



- L'étoile «*» est un quantifieur qui définit que ce qui précède le point d'interrogation peut exister zéro ou plusieurs fois.

Expression régulière	Texte accepté
/ab*c/	ac
	abc
	abbc
	abbbc
	...



- Le plus «+» est un quantifieur qui définit que ce qui précède le point d'interrogation peut exister **une** ou plusieurs fois.

Expression régulière	Texte accepté
/ab+c/	abc abbc abbbc ...



- Les expressions peuvent être combinées pour former des expressions plus complexes.

$/H(\ddot{a}|ae?)nde1/ = /H(ae?|\ddot{a})nde1/ = /H(a|ae|\ddot{a})nde1/$



- Le point « . » remplace n'importe quel caractère (sauf la fin de ligne `\n`).

Expression régulière	Texte accepté
/a.c/	abc
	aac
	axc
	...



- Les crochets «`[]`» permettent de spécifier un set de caractères possibles.

Expression régulière	Texte accepté
<code>[abc]</code>	a b c
<code>[a-z]</code>	une lettre en minuscule
<code>[0-9-]</code>	un chiffre ou le signe « <code>-</code> »
<code>[]0-9]</code>	un chiffre ou le signe « <code>]</code> »
<code>[]</code>	un espace



- Les crochets avec un circonflexe comme premier caractère «`[^]`» permettent de spécifier l'**inverse** d'un set de caractères possibles.

Expression régulière	Texte accepté
<code>[^ abc]</code>	d e f ...
<code>[^ a-z]</code>	un caractère sauf une lettre en minuscule



Les expressions régulières permettent de trouver un texte quelle que soit sa position. Par exemple `/lo/` laissera passer le texte «Hello»

- Le circonflexe «`^`» indique le **début** d'un texte.
- Le dollar «`$`» indique le **fin** d'un texte.

Expression régulière	Texte accepté
<code>^hello</code>	hello world
<code>der\$</code>	Darth Vader



- l'opération « $\{m, n\}$ » indique que ce qui précède doit être au minimum m fois présent et au maximum n fois. $\{m, \}$ signifie au minimum m fois, sans préciser de maximum.

Expression régulière	Texte accepté
$ab\{3, 5\}c$	abbbbc abbbbcb abbbbbcb

$/x?/ = /x\{0, 1\}/$

$/x+/ = /x\{1, \}/$

$/x*/ = /x\{0, \}/$



- Les *classes de caractères* sont représentées par «[:classe:]»

Classe	VIM	ASCII
[:alnum:]		[A-Za-z0-9]
[:word:]	\w	[A-Za-z0-9_]
	\W	[^A-Za-z0-9_]
[:alpha:]	\a	[A-Za-z]
[:blank:]	\s	[\t]
	\S	[^ \t\r\n\v\f]
[:digit:]	\d	[0-9]
	\D	[^0-9]
[:lower:]	\l	[a-z]
[:upper:]	\u	[A-Z]



Les classes doivent être utilisées dans un set de caractères.

Exemples : `/[[:alpha:]]/`, ou `/[[:digit:]]a-f/`.



Voici encore une liste d'éléments d'expressions régulières

Élément	Explication
<code>\b</code>	condition vraie à la limite d'un mot
<code>\t</code>	tabulateur
<code>\ooo</code>	code octal du caractère
<code>\xNN</code>	code hexadécimal du caractère
<code>\autre</code>	autre caractère considéré littéralement par exemple <code>\/. /</code>



Donnez les expressions régulières pour reconnaître les textes suivants :

- Les codes postaux suisses
- Les numéros de téléphone suisses
- Les noms de domaines
- Les adresses IPv4
- Les adresses IPv6
- Les adresses MAC
- Les adresses e-mail
- les URLs