

2-005-298-1

Bernard Desgraupes

Introduction aux expressions régulières



Table des matières

Introduction	XV
Historique des expressions régulières	XV
Champs d'application	XVII
Comment lire cet ouvrage	XVII
Le vocabulaire des expressions régulières	XIX
1 Expressions élémentaires	1
1.1 Généralités	1
1.2 La recherche avec les expressions régulières	3
Interprétation d'une expression régulière	6
1.3 Syntaxe étendue des expressions régulières	7
1.3.1 Métacaractères de la syntaxe étendue	7
1.3.2 Échappements de la syntaxe étendue	8
1.3.3 Autres quantificateurs	11
1.4 Avidité des recherches régulières	11
1.5 Le remplacement avec les expressions régulières	12
1.5.1 Métacaractères pour le remplacement	13
1.5.2 Séquences de contrôle	14
1.6 Contextes d'utilisation	15
Éditeurs de texte	15
Langages de scripts	16
Utilitaires de recherche	18
1.7 Récapitulation de la syntaxe élémentaire	18
2 Les expressions régulières par l'exemple	21
2.1 Motifs de recherche	22
Exemple 1 Expressions usuelles	22
Exemple 2 Voyelles et consonnes	23
Exemple 3 Crochets	24
Exemple 4 Valeurs numériques	24
Exemple 5 Nombres décimaux	25
Exemple 6 Sections d'un document L ^A T _E X	26
Exemple 7 Balises dans un document HTML	26

Exemple 8	Adresses électroniques	27
Exemple 9	Chaînes entre guillemets	28
Exemple 10	Commentaires de fichiers source	29
Exemple 11	Sections d'un document HTML	33
2.2	Motifs de remplacement	34
Exemple 12	Transformation de données	35
Exemple 13	Création de sigles	35
Exemple 14	Traduction de code	36
Exemple 15	Extraire des URL	37
Exemple 16	Créer des ancres et des liens dans un document HTML . .	38
Exemple 17	Extraire des informations	39
Exemple 18	Reformatage d'une base de données	40
Exemple 19	Traitement dichotomique	41
3	Expressions avancées	43
3.1	Quantificateurs de la syntaxe avancée	43
3.1.1	Itérations	43
3.1.2	Quantificateurs minimaux	44
3.2	Parenthèses de la syntaxe avancée	45
3.2.1	Commentaires	45
3.2.2	Motifs non capturants	46
3.2.3	Motifs prospectifs	46
3.2.4	Motifs rétrospectifs	47
3.3	Échappements de la syntaxe avancée	48
3.3.1	Nouvelle interprétation de la contre-oblique	48
3.3.2	Séquences d'entrée de caractères	49
3.3.3	Raccourcis de classes	50
3.3.4	Séquences de positionnement	51
3.3.5	Séquences de référence arrière	53
3.3.6	Séquences de conversion	54
3.4	Crochets de la syntaxe avancée	55
3.4.1	Classes prédéfinies	56
3.4.2	Éléments de comparaison	58
3.4.3	Classes d'équivalence	59
3.5	Utilisation de la syntaxe avancée	60
Exemple 1	Lettres et mots doublés	60
Exemple 2	Adresse IP	61
Exemple 3	Nettoyage et suppression de balises	61
Exemple 4	Modification de casse	62
3.6	Points subtils	63
3.6.1	Efficacité des expressions régulières	63
3.6.2	Questions de sécurité	65
3.7	Récapitulation de la syntaxe avancée	67

4 Les expressions régulières avec Tcl	69
4.1 Les langages interprétatifs	69
4.2 Généralités	70
4.3 Tcl-ismes	72
4.3.1 Syntaxe usuelle	72
4.3.2 Syntaxe particulière	73
Directives optionnelles	75
Sélecteur d'interprétation	77
4.4 Utilisation des expressions régulières avec Tcl	78
4.4.1 Fonctions Tcl pour la recherche et la substitution	78
La fonction <i>regexp</i>	78
Variables capturantes	79
Commutateurs	80
La fonction <i>regsub</i>	81
Commutateurs	82
Stockage des motifs	82
4.4.2 Exemples	83
Exemple 1 Décompte d'objets	83
Exemple 2 Recherche par indices	84
Exemple 3 L'option <i>inline</i>	84
Exemple 4 Simulation de classes d'équivalence	85
Exemple 5 Expressions régulières longues	85
5 Les expressions régulières avec Perl	87
5.1 Généralités	87
5.1.1 Support Unicode	87
L'encodage UTF-8	88
Expressions régulières et Unicode	89
5.1.2 Syntaxe avancée avec Perl	90
5.2 Perlismes	91
5.2.1 Séquences d'échappement Unicode	91
5.2.2 Directives de compilation	96
5.2.3 Motifs particuliers	98
Motifs d'évaluation	99
Motifs conditionnels	100
Motifs autonomes	100
Motifs post-évalués	101
5.2.4 Motifs de substitution	102
Variables des motifs de remplacement	102
Séquences de conversion	103
5.2.5 Idiosyncrasies	104
Ancres	104
Négation de classes	105
Motifs non capturants	105
5.3 Utilisation des expressions régulières avec Perl	106

5.3.1	L'opérateur <code>m/ /</code>	106
5.3.2	L'opérateur <code>s/ / /</code>	108
5.3.3	L'opérateur <code>qr/ /</code>	109
5.4	Exemples d'utilisation	110
Exemple 1	Expressions multiples	110
Exemple 2	Évaluations	111
Exemple 3	Fréquence des mots	111
Exemple 4	Analyseur lexical	111
Exemple 5	Sous-motifs maximaux	112
6	Les expressions régulières avec Python	113
6.1	Généralités	113
6.2	Pythonismes	114
	Les chaînes brutes	114
	Les motifs nominatifs	115
	Les motifs de substitution	116
	Les directives de compilation	117
6.3	Les expressions régulières orientées objet	117
	Syntaxe objet	118
	Syntaxe globale	119
6.3.1	Les méthodes de la classe <i>RegexObject</i>	119
	Attributs de la classe <i>RegexObject</i>	122
6.3.2	Les fonctions du module <i>re</i> de Python	123
	Drapeaux du module <i>re</i>	124
6.3.3	Les méthodes de la classe <i>MatchObject</i>	125
	Attributs de la classe <i>MatchObject</i>	128
6.3.4	Syntaxe comparée des modules <i>regex</i> et <i>re</i>	128
	Différences entre les modules	129
	Motifs de recherche	129
	Motifs de substitution	130
6.3.5	Analyse du premier caractère	130
6.4	Exemples	131
Exemple 1	Identifier un numéro de version	131
Exemple 2	Formater des nombres en virgule flottante	131
Exemple 3	Scinder une chaîne entre guillemets	132
7	Les expressions régulières avec Awk	135
7.1	Généralités	135
7.2	Awkismes	136
	Les options globales de la ligne de commande	139
7.3	Utilisation des expressions régulières avec Awk	140
7.3.1	Les opérateurs de correspondance	140
7.3.2	Expressions constantes ou interpolées	141
7.3.3	Les séparateurs	142
	Séparateurs d'enregistrements	142

	Séparateur de champs	143
7.3.4	Les fonctions de recherche d'Awk	144
7.3.5	Majuscules et minuscules	146
7.4	Exemples	147
	Exemple 1 Expressions régulières multiples	147
	Exemple 2 Extraction de données	148
	Exemple 3 Fréquence de mots	148
8	Expressions régulières avec JavaScript	149
8.1	Généralités	149
8.2	Syntaxe disponible	150
8.3	Utilisation des expressions avec JavaScript	152
	8.3.1 Mode de fonctionnement	152
	Le problème de la double interprétation	154
	8.3.2 Classes JavaScript	154
	8.3.3 Méthodes	154
	8.3.4 Options globales de recherche	160
	8.3.5 Propriétés	161
8.4	Exemples de scripts	163
	Exemple 1 Analyse de données interactives	163
	Exemple 2 Conversion Celsius-Fahrenheit	164
	Exemple 3 Tableaux de données	165
9	Outils de recherche	167
9.1	La famille grep	167
	Les options	168
	Les fichiers	168
9.2	Différences entre grep, egrep et fgrep	170
9.3	Exemples de recherche régulière avec egrep	172
	Exemple 1 Fichiers cachés	172
	Exemple 2 Extraction de données	173
	Exemple 3 Utilisation des options	173
10	Les éditeurs de texte	175
10.1	L'éditeur sed	176
	10.1.1 Syntaxe de base de sed	176
	10.1.2 Versions avancées de sed	179
	10.1.3 Exemples d'expressions régulières avec sed	182
	Exemple 1 Formatage de nombres	182
	Exemple 2 Suppression de doublons	182
	Exemple 3 Conversion de balises	183
10.2	L'éditeur emacs	183
	10.2.1 Les expressions régulières avec emacs	184
	La syntaxe	184
	Les classes modales	185

10.2.2	Optimisation selon le premier caractère	186
10.2.3	Les fonctions de recherche de emacs	186
	Recherche dans le tampon	187
	Recherche dans une chaîne	188
	Recherche incrémentale	188
	Remplacement	189
	Récupération des résultats	190
	Fonctions de type POSIX	191
10.3	L'éditeur Alpha	192
	Expressions régulières avec Alpha	192
10.3.1	Fonctions de recherche dans Alpha	193
10.3.2	Filtrage avec Alpha	195
A	Référence rapide	197
	Syntaxe formelle	198
	Fonctionnalités supportées	199
	Tableau synoptique des séquences d'échappement	200
	Classes de la syntaxe POSIX	202
	Directives de Tcl	202
	Propriétés de type <i>In</i> des caractères Unicode	203
	Propriétés de type <i>Is</i> des caractères Unicode	205
	Catégories Unicode générales de Perl	207
B	Algorithmes et fonctionnement interne	209
B.1	Modélisation des langages	210
B.1.1	Alphabets, mots et langages	210
B.1.2	Automates finis	211
B.1.3	Expressions régulières et automates	212
B.2	Moteurs de recherche régulière	213
B.2.1	Moteurs déterministes et non déterministes	213
B.2.2	Rétrogradation et rétrocession	214
	Tests de rapidité	215
B.3	Optimisation des algorithmes	216
B.4	Code source des expressions régulières	217
C	Adresses internet	219
D	Lexique anglais-français	223
	Index	225