

Introdução às

Expressões Regulares

Michael Fitzgerald

Novatec

Authorized Portuguese translation of the English edition of titled *Introducing Regular Expressions, First Edition* ISBN 9781449392680 © 2012 Michael Fitzgerald. This translation is published and sold by permission of O'Reilly Media, Inc., the owner of all rights to publish and sell the same.

Tradução em português autorizada da edição em inglês da obra *Introducing Regular Expressions, First Edition* ISBN 9781449392680 © 2012 Michael Fitzgerald. Esta tradução é publicada e vendida com a permissão da O'Reilly Media, Inc., detentora de todos os direitos para publicação e venda desta obra.

© Novatec Editora Ltda. [2012].

Todos os direitos reservados e protegidos pela Lei 9610 de 19/02/1998. É proibida a reprodução desta obra, mesmo parcial, por qualquer processo, sem prévia autorização, por escrito, do autor e da Editora.

Editor: Rubens Prates

Tradução: Lúcia Ayako Kinoshita

Revisão gramatical: Marta Almeida de Sá

Editoração eletrônica: Carolina Kuwabata

ISBN: 978-85-7522-330-7

Histórico de impressões:

Setembro/2012 Primeira edição

Novatec Editora Ltda.

Rua Luís Antônio dos Santos 110

02460-000 – São Paulo, SP – Brasil

Tel.: +55 11 2959-6529

Fax: +55 11 2950-8869

E-mail: novatec@novatec.com.br

Site: www.novatec.com.br

Twitter: twitter.com/novateceditora

Facebook: facebook.com/novatec

LinkedIn: linkedin.com/in/novatec

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Fitzgerald, Michael
Introdução às expressões regulares / Michael
Fitzgerald ; [tradução Lúcia Ayako Kinoshita]. --
São Paulo : Novatec Editora ; Sebastopol, CA :
O'Reilly, 2012.

Título original: *Introducing regular
expressions*
ISBN 978-85-7522-330-7 (Novatec)

1. Expressões regulares I. Título.

12-11222

CDD-005.115

Índices para catálogo sistemático:

1. Expressões regulares : Ciência da computação
005.115
MP20120905

Sumário

Prefácio	9
A quem se destina este livro	11
O que você precisa para usar este livro	12
Convenções usadas neste livro	12
Uso de exemplos de código de acordo com a política da O'Reilly.....	13
Como entrar em contato conosco	13
Agradecimentos	14
 Capítulo 1 ■ O que é uma Expressão Regular?	15
Conhecendo o Regexpal	16
Correspondendo a um número de telefone no padrão norte-americano.....	18
Correspondendo a dígitos usando uma classe de caracteres	19
Usando shorthand de caracteres	20
Correspondendo a qualquer caractere.....	21
Grupos de captura e referências para trás	21
Usando quantificadores.....	22
Usando literais com escape.....	24
Uma amostra de aplicativos.....	25
O que você aprendeu no Capítulo 1	27
Notas técnicas.....	28
 Capítulo 2 ■ Correspondência de padrões simples.....	30
Correspondendo a strings literais	32
Correspondendo a dígitos	32
Correspondendo a não-dígitos.....	33
Correspondendo a caracteres de palavra e não-caracteres de palavra.....	34
Correspondendo a espaços em branco	36
Correspondendo a qualquer caractere novamente.....	37
Inserindo marcação no texto.....	39
Usando sed para marcação de texto	40

Usando Perl para marcação de texto.....	42
O que você aprendeu no Capítulo 2.....	43
Notas técnicas.....	44
Capítulo 3 ■ Bordas	47
Início e fim de uma linha.....	47
Bordas de palavra e não-bordas de palavra	49
Outras âncoras	51
Especificando um grupo de caracteres como literais	52
Adicionando tags	53
Adicionado tags com o sed	54
Adicionando tags usando Perl	55
O que você aprendeu no Capítulo 3.....	57
Notas técnicas.....	57
Capítulo 4 ■ Alternância, grupos e retrovisores.....	59
Alternância	59
Subpadrões	63
Capturando grupos e usando retrovisores	65
Grupos nomeados	67
Grupos de não-captura	68
Grupos atômicos	69
O que você aprendeu no Capítulo 4.....	70
Notas técnicas.....	70
Capítulo 5 ■ Classes de caracteres	72
Classes de caracteres negados	74
União e diferença	74
Classes de caracteres POSIX	75
O que você aprendeu no Capítulo 5.....	77
Notas técnicas.....	77
Capítulo 6 ■ Correspondendo a caracteres Unicode e outros caracteres.....	79
Correspondendo a um caractere Unicode.....	80
Usando vim	82
Correspondendo a caracteres com números octais.....	82
Correspondendo a propriedades de caracteres Unicode	83
Correspondendo a caracteres de controle	86

O que você aprendeu no Capítulo 6.....	88
Notas técnicas.....	89
Capítulo 7 ■ Quantificadores.....	91
Guloso, preguiçoso e possessivo.....	91
Correspondendo por meio de *, + e ?	92
Efetuando a correspondência um número específico de vezes.....	94
Quantificadores preguiçosos.....	95
Quantificadores possessivos	97
O que você aprendeu no Capítulo 7.....	98
Notas técnicas.....	98
Capítulo 8 ■ Lookarounds	99
Lookaheads positivos	99
Lookaheads negativos	101
Lookbehinds positivos	102
Lookbehinds negativos.....	102
O que você aprendeu no Capítulo 8.....	103
Notas técnicas.....	104
Capítulo 9 ■ Inserindo marcação HTML em um documento	105
Correspondendo a tags	105
Transformando texto normal usando sed	107
Substituição usando sed.....	108
Lidando com algarismos romanos no sed	109
Lidando com um parágrafo específico usando sed.....	110
Lidando com as linhas do poema usando sed	110
Concatenando tags	111
Usando um arquivo de comandos com o sed.....	112
Transformando texto normal usando Perl.....	114
Lidando com algarismos romanos usando Perl	115
Lidando com um parágrafo específico usando Perl	116
Lidando com as linhas do poema usando Perl.....	116
Usando um arquivo de comandos com Perl.....	117
O que você aprendeu no Capítulo 9.....	119
Notas técnicas.....	119

Capítulo 10 ■ O fim do começo	121
Aprendendo mais	123
Ferramentas, implementações e bibliotecas interessantes	124
Perl	124
PCRE	124
Ruby (Oniguruma)	125
Python	125
RE2	125
Correspondendo a um número de telefone no padrão norte-americano...	126
Correspondendo a um endereço de email	127
O que você aprendeu no Capítulo 10	127
 Apêndice ■ Referência para Expressões Regulares	128
Expressões regulares no QED	128
Metacaracteres	129
Shorthand de caracteres	130
Espaços em branco	131
Caracteres Unicode Brancos	131
Caracteres de controle	133
Propriedades de caracteres	134
Nomes de script para propriedades de caracteres	135
Classes de Caracteres POSIX	136
Opções/modificadores	137
Tabela de códigos ASCII com Regex	138
Notas técnicas	142
 Glossário de expressões regulares	143
 Sobre o autor	153
 Colofão	154
 Índice Remissivo	156