

Table of Contents

Preface xviii

1: Introduction to Regular Expressions 1

Solving Real Problems 2

Regular Expressions as a Language 4

The Filename Analogy 4

The Language Analogy 5

The Regular-Expression Frame of Mind 6

If You Have Some Regular-Expression Experience 6

Searching Text Files: Egrep 6

Egrep Metacharacters 8

Start and End of the Line 8

Character Classes 9

Matching Any Character with Dot 11

Alternation 13

Ignoring Differences in Capitalization 14

Word Boundaries 15

In a Nutsell 16

Optional Items 17

Other Quantifiers: Repetition 18

Parentheses and Backreferences 20

The Great Escape 22

Expanding the Foundation 23

Linguistic Diversification 23

The Goal of a Regular Expression 23

A Few More Examples 23

Regular Expression Nonclature 27

Improving on the Status Quo 30

Summary 32

Personal Glimpses 33

2: Extended Introductory Examples 35

About the Examples 36

A Short Introduction to Perl 37

Matching Text with Regular Expressions 38

Toward a More Real-World Example 40

Side Effects of a Successful Match 40

Intertwined Regular Expressions 43

Intermission 49

Modifying Text with Regular Expressions 50

Example: Form Letter 50

Example: Prettyfying a Stock Price 51

Automated Editing 53

A Small Mail Utility 53

Adding Commas to a Number with Lookaround 59

Text-to-HTML Conversion 67

That Doubled-Word Thing 77

3: Overview of Regular Expression Features and Flavors 83

A Casual Stroll Across the Regex Landscape 85

The Origins of Regular Expressions 85

At a Glance 91

Care and Handling of Regular Expressions 93

Integrated Handling 94

Procedural and Object-Oriented Handling 95

A Search-and-Replace Example 98

Search and Replace in Other Languages 100

Care and Handling: Summary 101

Strings, Character Encodings, and Modes 101

Strings as Regular Expressions 101

Character-Encoding Issues 105

Unicode 106

Regex Modes and Match Modes 110

Common Metacharacters and Features 113

Character Representations	115
Character Classes and Class-Like Constructs	118
Anchor and Other "Zero-Width Assertions"	129
Comments and Mode Modifiers	135
Grouping, Capturing, Conditionals, and Control	137
Guide to the Advanced Chapters	142

4. *The Mechanics of Expression Processing* 143

Start Your Engines!	143
Two Kinds of Engines	144
New Standards	144
Regex Engine Types	145
From the Department of Redundancy Department	146
Testing the Engine Type	146
Match Basics	147
About the Examples	147
Rule 1: The Match That Begins Earliest Wins	148
Engine Pieces and Parts	149
Rule 2: The Standard Quantifiers Are Greedy	151
Regex-Directed Versus Text-Directed	153
NFA Engine: Regex-Directed	153
DFA Engine: Text-Directed	155
First Thoughts: NFA and DFA in Comparison	156
Backtracking	157
A Really Crummy Analogy	158
Two Important Points on Backtracking	159
Saved States	159
Backtracking and Greediness	162
More About Greediness and Backtracking	163
Problems of Greediness	164
Multi-Character "Quotes"	165
Using Lazy Quantifiers	166
Greediness and Laziness Always Favor a Match	167
The Essence of Greediness, Laziness, and Backtracking	168
Possessive Quantifiers and Atomic Grouping	169
Possessive Quantifiers, <code>?</code> , <code>++</code> , <code>++</code> , and <code>[m,n]++</code>	172
The Backtracking of Lookaround	173
Is Alternation Greedy?	174
Taking Advantage of Ordered Alternation	175