

Table of Contents

- About the Authorix**
- About the Technical Reviewerxi**
- Acknowledgmentsxiii**

- Chapter 1: What are Regexes and Grammars? 1**
 - 1.1 Use Cases 1
 - Searching 1
 - Validation.....3
 - Parsing3
 - 1.2 Regexes or Regular Expressions?.....4
 - 1.3 What’s So Special about Perl 6 Regexes?.....5

- Chapter 2: Getting Started with Perl 6.....7**
 - 2.1 Installing Rakudo Perl 67
 - Rakudo Star from Native Installers.....8
 - Binary Linux Packages8
 - Docker-Based Installation.....8
 - 2.2 Using Rakudo Perl 6.....9
 - 2.3 Obtaining the Code Examples 11
 - 2.4 First Steps with Perl 6..... 11
 - Variables and Values 12
 - Strings 13

TABLE OF CONTENTS

Control Structures 14

Functions, Classes, and Methods 15

Learning More About Perl 6 16

2.5 Summary..... 16

Chapter 3: The Building Blocks of Regexes 17

3.1 Literals 17

3.2 Meta Characters vs. Literals 18

3.3 Anchors 19

3.4 Predefined Character Classes 21

 User-Defined Character Classes 23

 Unicode Properties 25

3.5 Quantifiers..... 26

 Greedy and Frugal Quantifiers 27

 Quantifiers with Separators..... 27

3.6 Disjunction 28

3.7 Conjunction 29

3.8 Zero-Width Assertions..... 30

3.9 Summary..... 32

Chapter 4: Regexes and Perl 6 Code..... 33

4.1 Smart-Matching 33

4.2 Quote Forms..... 34

4.3 Modifiers 35

4.4 Comb and Split..... 38

4.5 Substitution..... 39

4.6 Crossing the Code and Regex Boundary 42

4.7 Summary..... 46

Chapter 5: Extracting Data from Regex Matches	47
5.1 Positional Captures	47
5.2 The Match Object	48
Nesting of Captures	49
Quantified Captures	49
5.3 Named Captures	50
5.4 Backreferences	52
Excursion: Primality Test with Backreferences.....	52
5.5 Match Objects Revisited	55
5.6 Summary.....	56
Chapter 6: Regex Mechanics	57
6.1 Matching with State Machines	57
Deterministic State Machines	57
Nondeterministic State Machines	62
6.2 Regex Control Flow	64
6.3 Backtracking	65
6.4 Why Would You Want to Avoid Backtracking?	68
Performance	68
Correctness	70
6.5 Frugal Quantifiers and Backtracking	71
6.6 Longest Token Matching	71
6.7 Summary.....	73
Chapter 7: Regex Techniques	75
7.1 Know Your Data Format	75
Well-Defined Data Formats	75
Exploring Data Formats	76
7.2 Think About Invalid Inputs.....	78

TABLE OF CONTENTS

7.3 Use Anchors	79
7.4 Matching Quoted Strings	80
Quoted Strings with Escaping Sequences.....	81
7.5 Testing Regexes	83
7.6 Summary.....	89
Chapter 8: Reusing and Composing Regexes	91
8.1 Named Regexes	91
Lexical Analysis and Backtracking Control.....	93
8.2 Whitespace.....	95
8.3 Grammars	98
8.4 Code Reuse with Grammars	100
Inheritance	100
Role Composition.....	102
8.5 Proto Regexes	104
8.6 Summary.....	108
Chapter 9: Parsing with Grammars	109
9.1 Understanding Grammars	109
Recursive Descent Parsing and Precedence	112
Left Recursion and Other Traps	113
9.2 Starting Simple	114
9.3 Assembling Complete Grammars.....	116
9.4 Debugging Grammars	116
9.5 Parsing Whitespace and Comments	121
9.6 Keeping State.....	123
Implementing Lexical Scoping with Dynamic Variables	128
Scoping Through Explicit Symbol Tables	131
9.7 Summary.....	134

Chapter 10: Extracting Data from Matches	135
10.1 Action Objects	136
10.2 Building ASTs with Action Objects	141
10.3 Keeping State in Action Objects.....	143
10.4 Summary.....	145
Chapter 11: Generating Good Parse Error Messages.....	147
11.1 Exploring the Problem.....	147
11.2 Assertions	149
11.3 Improved Position Reporting.....	151
11.4 High-Water Marks	153
11.5 Parser Combinator and FAILGOAL.....	155
11.6 Which Techniques to Use?	157
11.7 Summary.....	158
Chapter 12: Unicode and Natural Language	159
12.1 Writing Systems	159
12.2 Bytes, Code Points, Graphemes, and Glyphs.....	161
Grapheme Clusters	161
Glyphs.....	163
12.3 Unicode Properties.....	163
12.4 Summary.....	164
Chapter 13: Case Studies.....	165
13.1 S-Expressions	165
Parsing S-Expressions.....	166
Data Extraction	171
13.2 Mathematical Expressions and Operator Precedence Parsers	174
A Simple Operator Precedence Parser	174
A More Flexible Approach.....	180

TABLE OF CONTENTS

13.3 Pythonesque, an Indentation-Based Language..... 183

 A Grammar for Pythonesque 184

 Action Objects 190

13.4 Summary..... 193

Index..... 195