



From *Introduction to Regular Expressions in SAS®*.
Full book available for purchase [here](#).

Contents

About This Book	vii
About The Author	xi
Acknowledgments	xiii
Chapter 1: Introduction	1
1.1 Purpose of This Book	1
1.2 Layout of This Book	1
1.3 Defining Regular Expressions.....	2
1.4 Motivational Examples	3
1.4.1 Extract, Transform, and Load (ETL).....	3
1.4.2 Data Manipulation	4
1.4.3 Data Enrichment	5
Chapter 2: Getting Started with Regular Expressions	9
2.1 Introduction	10
2.1.1 RegEx Test Code	11
2.2 Special Characters.....	13
2.3 Basic Metacharacters.....	15
2.3.1 Wildcard.....	15
2.3.2 Word.....	15
2.3.3 Non-word.....	16
2.3.4 Tab.....	16
2.3.5 Whitespace.....	17
2.3.6 Non-whitespace.....	17
2.3.7 Digit.....	17
2.3.8 Non-digit.....	18
2.3.9 Newline	18
2.3.10 Bell	19

2.3.11 Control Character	20
2.3.12 Octal	20
2.3.13 Hexadecimal	21
2.4 Character Classes	21
2.4.1 List	21
2.4.2 Not List	22
2.4.3 Range	22
2.5 Modifiers	23
2.5.1 Case Modifiers	23
2.5.2 Repetition Modifiers	25
2.6 Options	32
2.6.1 Ignore Case	32
2.6.2 Single Line	32
2.6.3 Multiline	33
2.6.4 Compile Once	33
2.6.5 Substitution Operator	34
2.7 Zero-width Metacharacters	34
2.7.1 Start of Line	35
2.7.2 End of Line	35
2.7.3 Word Boundary	35
2.7.4 Non-word Boundary	36
2.7.5 String Start	36
2.8 Summary	37
Chapter 3: Using Regular Expressions in SAS.....	39
3.1 Introduction	39
3.1.1 Capture Buffer	39
3.2 Built-in SAS Functions	40
3.2.1 PRXPARSE	40
3.2.2 PRXMATCH	42
3.2.3 PRXCHANGE	43
3.2.4 PRXPOSN	46
3.2.5 PRXPAREN	47

3.3 Built-in SAS Call Routines	49
3.3.1 CALL PRXCHANGE	50
3.3.2 CALL PRXPOSN	54
3.3.3 CALL PRXSUBSTR	56
3.3.4 CALL PRXNEXT	57
3.3.5 CALL PRXDEBUG	59
3.3.6 CALL PRXFREE.....	62
3.4 Summary	63
Chapter 4: Applications of Regular Expressions in SAS	65
4.1 Introduction	65
4.1.1 Random PII Generator	66
4.2 Data Cleansing and Standardization.....	72
4.3 Information Extraction	77
4.4 Search and Replacement	80
4.5 Summary	83
4.5.1 Start Small	83
4.5.2 Think Big.....	83
Appendix A: Perl Version Notes	85
Appendix B: ASCII Code Lookup Tables	87
Non-Printing Characters	87
Printing Characters	89
Appendix C: POSIX Metacharacters	97
Index	101