# Contents

**Preface** . . . . . *page* ix

**Chapter 1** **Algorithms on Words** . . . . . 1

1.0 Introduction . . . . . 1
1.1 Words . . . . . 3
1.2 Elementary algorithms . . . . . 7
1.3 Tries and automata . . . . . 15
1.4 Pattern matching . . . . . 36
1.5 Transducers . . . . . 40
1.6 Parsing . . . . . 52
1.7 Word enumeration . . . . . 69
1.8 Probability distributions on words . . . . . 74
1.9 Statistics on words . . . . . 91
Problems . . . . . 96
Notes . . . . . 101

**Chapter 2** **Structures for Indexes** . . . . . 106

2.0 Introduction . . . . . 106
2.1 Suffix trie . . . . . 107
2.2 Suffix tree . . . . . 113
2.3 Contexts of factors . . . . . 121
2.4 Suffix automaton . . . . . 127
2.5 Compact suffix automaton . . . . . 138
2.6 Indexes . . . . . 141
2.7 Finding regularities . . . . . 150
2.8 Pattern matching machine . . . . . 155
Problems . . . . . 160
Notes . . . . . 161

**Chapter 3 Symbolic Natural Language Processing** . . . . . 164

3.0 Introduction . . . . . 164
3.1 From letters to words . . . . . 165
3.2 From words to sentences . . . . . 199
Notes . . . . . 208

**Chapter 4 Statistical Natural Language Processing** . . . . . 210

4.0 Introduction . . . . . 210
4.1 Preliminaries . . . . . 211
4.2 Algorithms . . . . . 213
4.3 Application to speech recognition . . . . . 226
Notes . . . . . 239

**Chapter 5 Inference of Network Expressions** . . . . . 241

5.0 Introduction . . . . . 241
5.1 Inferring simple network expressions: models and related problems . . . . . 242
5.2 Algorithms . . . . . 248
5.3 Inferring network expressions with spacers . . . . . 256
5.4 Related issues . . . . . 260
5.5 Open problems . . . . . 264
Notes . . . . . 265

**Chapter 6 Statistics on Words with Applications to Biological Sequences** . . . . . 268

6.0 Introduction . . . . . 268
6.1 Probabilistic models for biological sequences . . . . . 270
6.2 Overlapping and nonoverlapping occurrences . . . . . 277
6.3 Word locations along a sequence . . . . . 281
6.4 Word count distribution . . . . . 289
6.5 Renewal count distribution . . . . . 311
6.6 Occurrences and counts of multiple patterns . . . . . 315
6.7 Some applications to DNA sequences . . . . . 328
6.8 Some probabilistic and statistical tools . . . . . 338
Notes . . . . . 346

**Chapter 7 Analytic Approach to Pattern Matching** . . . . . 353

7.0 Introduction . . . . . 353
7.1 Probabilistic models . . . . . 356

7.2 Exact string matching . . . . . 359
7.3 Generalized string matching . . . . . 375
7.4 Subsequence pattern matching . . . . . 393
7.5 Generalized subsequence problem . . . . . 407
7.6 Self-repetitive pattern matching . . . . . 413
Problems . . . . . 425
Notes . . . . . 427

**Chapter 8 Periodic Structures in Words** . . . . . 430

8.0 Introduction . . . . . 430
8.1 Definitions and preliminary results . . . . . 431
8.2 Counting maximal repetitions . . . . . 433
8.3 Basic algorithmic tools . . . . . 439
8.4 Finding all maximal repetitions in a word . . . . . 443
8.5 Finding quasi-squares in two words . . . . . 448
8.6 Finding repeats with a fixed gap . . . . . 450
8.7 Computing local periods of a word . . . . . 454
8.8 Finding approximate repetitions . . . . . 461
Notes . . . . . 474

**Chapter 9 Counting, Coding, and Sampling with Words** . . . . . 478

9.0 Introduction . . . . . 478
9.1 Counting: walks in sectors of the plane . . . . . 480
9.2 Sampling: polygons, animals, and polyominoes . . . . . 492
9.3 Coding: trees and maps . . . . . 504
Problems . . . . . 516
Notes . . . . . 518

**Chapter 10 Words in Number Theory** . . . . . 520

10.0 Introduction . . . . . 520
10.1 Morphic and automatic sequences: definitions and generalities . . . . . 521
10.2 $d$-Kernels and properties of automatic sequences . . . . . 526
10.3 Christol's algebraic characterization of automatic sequences . . . . . 536
10.4 An application to transcendence in positive characteristic . . . . . 541
10.5 An application to transcendental power series over the rationals . . . . . 543
10.6 An application to transcendence of real numbers . . . . . 544

10.7 The Tribonacci word . . . . . 546
10.8 The Rauzy fractal . . . . . 552
10.9 An application to simultaneous approximation . . . . . 564
Problems . . . . . 567
Notes . . . . . 574

**References** . . . . . 579

**General Index** . . . . . 603