

Computing Patterns in Strings

Bill Smyth



Contents

<i>Preface</i>	<i>ix</i>
Part I Strings and Algorithms	1
Chapter 1 Properties of Strings	3
1.1 Strings of Pearls	3
1.2 Linear Strings	5
1.3 Periodicity	14
1.4 Necklaces	25
Chapter 2 Patterns? What Patterns?	35
2.1 Intrinsic Patterns (Part II)	35
2.2 Specific Patterns (Part III)	41
2.3 Generic Patterns (Part IV)	51
Chapter 3 Strings Famous and Infamous	61
3.1 Avoidance Problems and Morphisms	61
3.2 Thue Strings (2, 3)	65
3.3 Thue Strings (3, 2)	72
3.4 Fibostings (2, 4)	76
Chapter 4 Good Algorithms and Good Test Data	89
4.1 Good Algorithms	89
4.2 Distinct Patterns	94
4.3 Distinct Borders	100

Part II	Computing Intrinsic Patterns	109
Chapter 5	Trees Derived from Strings	111
5.1	Border Trees	111
5.2	Suffix Trees	113
5.2.1	Preliminaries	114
5.2.2	McCreight's Algorithm	117
5.2.3	Ukkonen's Algorithm	121
5.2.4	Farach's Algorithm	126
5.2.5	Application and Implementation	137
5.3	Alternative Suffix-Based Structures	140
5.3.1	Directed Acyclic Word Graphs	140
5.3.2	Suffix Arrays — Saving the Best till Last?	149
Chapter 6	Decomposing a String	157
6.1	Lyndon Decomposition: Duval's Algorithm	158
6.2	Lyndon Applications	167
6.3	s -Factorization: Lempel-Ziv	175
Part III	Computing Specific Patterns	179
Chapter 7	Basic Algorithms	181
7.1	Knuth-Morris-Pratt	181
7.2	Boyer-Moore	187
7.3	Karp-Rabin	198
7.4	Dömölki-(Baeza-Yates)-Gonnet	202
7.5	Summary	206
Chapter 8	Son of BM Rides Again!	207
8.1	The BM Skip Loop	208
8.2	BM-Horspool	210
8.3	Frequency Considerations and BM-Sunday	212
8.4	BM-Galil	219
8.5	Turbo-BM	223
8.6	Daughter of KMP Rides Too!	226
8.7	Mix Your Own Algorithm	231
8.8	The Exact Complexity of Exact Pattern-Matching	233
Chapter 9	String Distance Algorithms	237
9.1	The Basic Recurrence	238
9.2	Wagner-Fischer <i>et al.</i>	241
9.3	Hirschberg	244
9.4	Hunt-Szymanski	250

9.5	Ukkonen-Myers	256
9.6	Summary	263
Chapter 10	Approximate Pattern-Matching	265
10.1	A General Distance-Based Algorithm	266
10.2	An Algorithm for k -Mismatches	269
10.3	Algorithms for k -Differences	274
10.3.1	Ukkonen's Algorithm	276
10.3.2	Myers' Algorithm	279
10.4	A Fast and Flexible Algorithm — Wu and Manber	286
10.5	The Complexity of Approximate Pattern-Matching	292
Chapter 11	Regular Expressions and Multiple Patterns	295
11.1	Regular Expression Algorithms	297
11.1.1	Non-Deterministic FA	297
11.1.2	Deterministic FA	302
11.1.3	Algorithm WM Revisited	305
11.2	Multiple Pattern Algorithms	309
11.2.1	Aho-Corasick FA: KMP Revisited	309
11.2.2	Commentz-Walter FA: BM Revisited	313
11.2.3	Approximate Patterns: WM Revisited Again!	315
11.2.4	Approximate Patterns: (Baeza-Yates)-Navarro	317
Part IV	Computing Generic Patterns	327
Chapter 12	Repetitions (Periodicity)	329
12.1	All Repetitions	331
12.1.1	Crochemore	331
12.1.2	Main and Lorentz	340
12.2	Runs	349
12.2.1	Leftmost Runs — Main	350
12.2.2	All Runs — Kolpakov and Kucherov	356
Chapter 13	Extensions of Periodicity	359
13.1	All Covers of a String — Algorithm LS	360
13.2	All Repeats — Algorithm FST	370
13.2.1	Computing the NE Tree	372
13.2.2	Computing the NE Array	375
13.3	k -Approximate Repeats — Schmidt	380
13.4	k -Approximate Periods — SIPS	396
	<i>Bibliography</i>	403
	<i>Index</i>	415