

Algorithms on Strings

MAXIME CROCHEMORE
Université de Marne-la-Vallée

CHRISTOPHE HANCART
Université de Rouen

THIERRY LECROQ
Université de Rouen



CAMBRIDGE
UNIVERSITY PRESS

Contents

<i>Preface</i>	<i>page vii</i>
1 Tools	1
1.1 Strings and automata	2
1.2 Some combinatorics	8
1.3 Algorithms and complexity	18
1.4 Implementation of automata	23
1.5 Basic pattern matching techniques	28
1.6 Borders and prefixes tables	40
2 Pattern matching automata	55
2.1 Trie of a dictionary	56
2.2 Searching for several strings	57
2.3 Implementation with failure function	65
2.4 Implementation with successor by default	72
2.5 Locating one string	82
2.6 Locating one string and failure function	85
2.7 Locating one string and successor by default	92
3 String searching with a sliding window	102
3.1 Searching without memory	103
3.2 Searching time	108
3.3 Computing the good suffix table	113
3.4 Automaton of the best factor	118
3.5 Searching with one memory	121
3.6 Searching with several memories	127
3.7 Dictionary searching	136
4 Suffix arrays	146
4.1 Searching a list of strings	147
4.2 Searching with the longest common prefixes	150

4.3	Preprocessing the list	155
4.4	Sorting suffixes	158
4.5	Sorting suffixes on bounded integer alphabets	164
4.6	Common prefixes of the suffixes	169
5	Structures for indexes	177
5.1	Suffix trie	178
5.2	Suffix tree	184
5.3	Contexts of factors	193
5.4	Suffix automaton	199
5.5	Compact suffix automaton	210
6	Indexes	219
6.1	Implementing an index	219
6.2	Basic operations	222
6.3	Transducer of positions	227
6.4	Repetitions	230
6.5	Forbidden strings	231
6.6	Search machine	234
6.7	Searching for conjugates	239
7	Alignments	243
7.1	Comparison of strings	244
7.2	Optimal alignment	251
7.3	Longest common subsequence	262
7.4	Alignment with gaps	273
7.5	Local alignment	276
7.6	Heuristic for local alignment	279
8	Approximate patterns	287
8.1	Approximate pattern matching with jokers	288
8.2	Approximate pattern matching with differences	293
8.3	Approximate pattern matching with mismatches	304
8.4	Approximate matching for short patterns	314
8.5	Heuristic for approximate pattern matching with differences	324
9	Local periods	332
9.1	Partitioning factors	332
9.2	Detection of powers	340
9.3	Detection of squares	345
9.4	Sorting suffixes	354
	<i>Bibliography</i>	364
	<i>Index</i>	377