

# Contents

	<i>Preface</i>	<i>page xi</i>
<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>A Warm-up</b>	<b>10</b>
	2.1 A Cubic-Time Algorithm	11
	2.2 A Quadratic-Time Algorithm	12
	2.3 A Linear-Time Algorithm	14
	2.4 Another Linear-Time Algorithm	16
	2.5 A Few Interesting Variants <sup>∞</sup>	17
<b>3</b>	<b>Random Sampling</b>	<b>23</b>
	3.1 Disk Model and Known Sequence Length	24
	3.2 Streaming Model and Known Sequence Length	26
	3.3 Streaming Model and Unknown Sequence Length	28
<b>4</b>	<b>List Ranking</b>	<b>32</b>
	4.1 The Pointer-Jumping Technique	33
	4.2 Parallel Algorithm Simulation in a Two-Level Memory	34
	4.3 A Divide-and-Conquer Approach	37
<b>5</b>	<b>Sorting Atomic Items</b>	<b>44</b>
	5.1 The Merge-Based Sorting Paradigm	45
	5.2 Lower Bounds	52
	5.3 The Distribution-Based Sorting Paradigm	57
	5.4 Sorting With Multi-Disks <sup>∞</sup>	67
<b>6</b>	<b>Set Intersection</b>	<b>72</b>
	6.1 Merge-Based Approach	74
	6.2 Mutual Partitioning	75

viii	<b>Contents</b>	
	6.3 Doubling Search	77
	6.4 Two-Level Storage Approach	79
<b>7</b>	<b>Sorting Strings</b>	82
	7.1 A Lower Bound	83
	7.2 RADIXSORT	84
	7.3 Multi-key QUICKSORT	90
	7.4 Some Observations on the Two-Level Memory Model <sup>∞</sup>	94
<b>8</b>	<b>The Dictionary Problem</b>	96
	8.1 Direct-Address Tables	97
	8.2 Hash Tables	98
	8.3 Universal Hashing	101
	8.4 A Simple (Static) Perfect Hash Table	106
	8.5 Cuckoo Hashing	111
	8.6 More on Static and Perfect Hashing: Minimal and Ordered	116
	8.7 Bloom Filters	121
<b>9</b>	<b>Searching Strings by Prefix</b>	128
	9.1 Array of String Pointers	129
	9.2 Locality-Preserving Front Coding <sup>∞</sup>	134
	9.3 Interpolation Search	136
	9.4 Compacted Trie	138
	9.5 Patricia Trie	142
	9.6 Managing Huge Dictionaries <sup>∞</sup>	145
<b>10</b>	<b>Searching Strings by Substring</b>	153
	10.1 Notation and Terminology	153
	10.2 The Suffix Array	154
	10.3 The Suffix Tree	172
	10.4 Some Interesting Problems	181
<b>11</b>	<b>Integer Coding</b>	194
	11.1 Elias Codes: $\gamma$ and $\delta$	197
	11.2 Rice Code	198
	11.3 PForDelta Code	199
	11.4 Variable-Byte Code and $(s, c)$ -Dense Codes	200
	11.5 Interpolative Code	203
	11.6 Elias–Fano Code	205
<b>12</b>	<b>Statistical Coding</b>	210
	12.1 Huffman Coding	210
	12.2 Arithmetic Coding	221
	12.3 Prediction by Partial Matching <sup>∞</sup>	234

	<b>Contents</b>	ix
<b>13</b>	<b>Dictionary-Based Compressors</b>	240
	13.1 LZ77	241
	13.2 LZ78	244
	13.3 LZW	246
	13.4 On the Optimality of Compressors <sup>∞</sup>	248
<b>14</b>	<b>Block-Sorting Compression</b>	252
	14.1 The Burrows–Wheeler Transform	253
	14.2 Two Other Simple Transforms	258
	14.3 The bzip Compressor	264
	14.4 On Compression Boosting <sup>∞</sup>	267
	14.5 On Compressed Indexing <sup>∞</sup>	268
<b>15</b>	<b>Compressed Data Structures</b>	274
	15.1 Compressed Representation of (Binary) Arrays	274
	15.2 Succinct Representation of Trees	284
	15.3 Succinct Representation of Graphs	291
<b>16</b>	<b>Conclusion</b>	299
	<i>Index</i>	302