

## Index

- algorithm
  - arithmetic coding, 221
  - binary MERGESORT, 45, 82
  - binary QUICKSORT, 57, 82
  - BWT compression, 252
  - CountingSort, 87
  - definition, 1
  - front coding, 131
  - GREEDSORT, 67
  - Huffman coding, 210
  - Huffman coding, canonical, 217
  - interpolation search, 136
  - list ranking, 32
  - locality-preserving front coding, 134
  - LZ77 compression, 189, 241
  - LZ78 compression, 244
  - LZW compression, 246
  - McCreight's suffix sorting, 177
  - multi-key QuickSort, 90
  - multi-way MERGESORT, 50
  - multi-way QUICKSORT, 65
  - permuting, 52
  - prediction by partial matching (PPM), 235
  - RADIXSORT, 84
  - range coding, 230
  - reservoir sampling, 28
  - scan-based suffix sort, 168
  - set intersection, 72, 126
  - skew or DC3 suffix sorting, 162
  - sorting, 35, 82
  - suffix sorting, 162, 168, 177
- approximate pattern matching, 181
- arithmetic coding, 221
- asymmetric numeral systems coding (ANS), 229
  
- binary integer code, 195
- bitonic sequence, 42
- Bloom filter, 121
  - compressed, 123
  - lower bound, 125
  - spectral, 124
- BOUNDEDQS., *see* QUICKSORT: bounding
  
- Brotli, 189, 240
- Burrows, Mike, 252
- Burrows–Wheeler transform (BWT), 252
  - backward transform, 255
  - bzip, 252, 264
  - compression booster, 267
  - FM-index, 268
  - forward transform, 253, 265
  - LF-mapping, 256, 270
- bzip, 252, 264
- bzip2, 250, 252, 265
  
- Cartesian tree, 185
- coin tossing
  - deterministic, 41
  - randomized, 40
- compressed data structure, 274
  - arbitrary graph, 295
  - arbitrary tree, 288
  - array, 274
  - binary tree, 284
  - Rank, 268, 275, 284, 286
  - Select, 207, 275, 280, 283, 286
  - Web graph, 292
- computational complexity, 3
  - polynomial vs. exponential, 4
  - golden rules, 6
  - I/O bottleneck, 3, 67
- Cormen, Thomas H., 3
- CountingSort, 87
- Cuckoo graph, 112
  
- data compression
  - arithmetic coding, 123, 221, 267
  - asymmetric numeral systems (ANS), 229
  - booster, 267
  - Brotli, 189, 240
  - bzip, 252, 264
  - compressed data structure, 274
  - entropy, 215, 216, 221, 228, 234
  - front coding, 131
  - graph encoding, 291

- gzip, 189, 243
- Huffman coding, 210, 261, 263
- Huffman coding, canonical, 217
- integer coding, 194
- $k^2$ -tree-based graph compression, 295
- locality-preserving front coding, 134
- LZ77, 189, 241
- LZ78, 244
- LZMA, 240
- LZW, 246
- Move-to-Front (MTF), 259
- optimality, 213, 221, 228, 248, 267
- prediction by partial matching (PPM), 235
- prefix-free code, 213
- Run Length Encoding (RLE), 263
- Shannon's coding theorem, 196, 215
- statistical coding, 210, 267
- tree encoding, 284
- Webgraph, 292
- data structure
  - Bloom filter, 121
  - Catesian tree, 185
  - compressed array, 274
  - compressed graph, 291
  - Cuckoo hashing, 111
  - FM-index, 124, 268
  - hash table, with chaining, 87, 98, 103
  - heap, 28, 49, 51
  - Patricia trie, 142
  - perfect hash table, 106
  - Rank, 269, 275, 283, 286
  - select, 207, 275, 279, 283, 286
  - String B-tree, 146
  - succinct tree encoding, 284, 288
  - suffix array, 154, 176, 191, 254, 266
  - suffix tree, 172, 190
  - ternary search tree, 93
  - trie, 84, 212, 245
  - trie, (un)compacted, 138
  - two-level indexing, 133
- dictionary
  - arrays and lists, 98
  - Cuckoo hashing, 111
  - dictionary compression, 240
  - direct access table, 97
  - hash table, with chaining, 98, 103, 115, 243
  - operations, 96
  - perfect hash table, 106
- dictionary-based compression
  - LZ77, 189, 241
  - LZ78, 244
  - LZss, 242
  - LZW, 246
  - optimality, 248
  - sliding window, 241
- Difference Cover modulo 3 (DC3), suffix sorting, 162
- direct access table, 97
- distinguishing (string) prefix, 83
- dyadic fraction, 222
- Elias- $\gamma$  and  $-\delta$  integer codes, 197, 293
- Elias-Fano integer code, 206
- entropy
  - 0-th order, 215, 221, 228, 261, 263
  - $k$ -th order, 216, 234, 248, 267, 270
- FM-index, 124, 268
- four Russians trick, 188
- front coding, 131
  - locality-preserving, 134, 145
- full-text indexing, 153, 175
- global rebuilding technique, 100, 116
- graph
  - compressed representation, 291
  - Cuckoo graph, 112
  - Elias-Fano encoding, 292
  - $k^2$ -tree-based compression, 295
  - random, 114, 117
  - Webgraph compressor, 293
- gzip, 189, 243
- hash function, 98
  - $d$ -choice hashing, 104, 111
  - (minimal ordered) perfect, 106, 116
  - simple uniform hashing, 100
  - universal hashing, 102, 105, 111
- hash table
  - $d$ -left hashing, 104
  - Bloom filter, 121
  - Cuckoo hashing, 111
  - dynamic, 111
  - hash function, 98
  - perfect and static, 106, 140
  - universal hashing, 102
  - with chaining, 87, 98, 103, 115
- Hollerith, Herman, 87
- Huffman coding, 210
  - canonical, 217
- Huffman, David A., 210
- I/Os
  - permuting lower bound, 52, 57
  - bottleneck, 67, 94
  - multi-disks, 67
  - pointer jumping, 36, 40
  - sequential vs. random, 7
  - sorting, 47, 52, 65, 94
  - sorting lower bound, 52, 55, 83
  - string prefix search, 130, 145, 148

- string substring-search, 156
- suffix sorting, 167, 172, 177, 266
- independent set
  - deterministic, 41
  - randomized, 40
- integer coding, 194
  - $(s, c)$ -dense, 201
  - binary, 195
  - Elias- $\gamma$  and  $-\delta$ , 197, 261, 263, 293
  - Elias-Fano, 206, 281, 292
  - gap coding, 195
  - interpolative, 203
  - optimality, 196
  - PForDelta, 199
  - Rice code, 198
  - statistical, 210
  - unary, 196
  - variable-byte, 200
- interpolation search, 136
- interpolative integer code, 203
  
- Jacobson, Guy, 285
  
- $k$ -mismatches problem, 181
- $k^2$ -ary tree, 295
- Knuth, Donald, 1, 48
  
- Lempel, Abraham, 189, 240
- locality-preserving front coding, 134, 145
- longest common prefix (LCP), 91, 143, 158, 182, 191
- lower bound
  - dictionary, 125
  - permuting, 52, 57
  - set intersection, 77
  - sorting items, 52, 55
  - sorting strings, 83, 94
- lowest common ancestor (LCA), 174, 184
  
- McCreight, Edward M., 177
- MERGESORT
  - binary, 45, 82
  - multi-way, 50
- model of computation
  - cache-oblivious model, 10
  - I/O bottleneck, 3
  - parallel (PRAM) model, 32
  - random access machine, 3
  - streaming model, 26
  - two-level memory model, 6, 35
- Move-to-Front compressor (MTF), 259
- multi-key QuickSort, 90
  
- Patricia trie, 142
- permuting, 52
- PForDelta integer code, 199
  
- pointer-jumping technique, 33
- pointerless programming, 274
- prediction by partial matching (PPM), 235
- problem
  - $k$ -mismatches, 181
  - approximate pattern matching, 181
  - data compression, 189, 210, 240, 252
  - dictionary, 96
  - full-text indexing, 153
  - GC-richness, 18
  - independent set, 40, 41
  - integer coding, 194
  - list ranking, 32
  - lowest common ancestor (LCA), 184
  - maximal subarray sum, 10
  - permuting, 44
  - range minimum query (RMQ), 181, 184
  - ranked item selection, 61
  - sampling, random, 23
  - set intersection, 72, 126
  - sorting items, 44
  - sorting strings, 82
  - string prefix search, 128
  - string substring-search, 153, 175
  - sum of integers, 5
  - text mining, 191
  - tree packing on disk, 149
  
- Qsort, 47, 64, 161
- QUICKSORT
  - binary, 57, 82
  - bounding recursive depth, 63
  - for strings, 90
  - multi-pivot selection, 65
  - multi-way, 65
  - pivot selection, 60, 91
  
- RADIXSORT
  - LSD-first, 84, 87
  - MSD-first, 84
- random sampling
  - known input length, 24, 26
  - unknown input length, 28
- RANDSELECT., *see* problem
- range coding, 230
- range minimum query (RMQ), 181, 184
- Rank data structure, 268, 275, 283, 286
- reservoir sampling, 28
- Rice integer code, 198
- Run Length Encoding compressor (RLE), 263
  
- $(s, c)$ -dense integer code, 201
- search engines, 72, 195, 200, 203
- SELECT data structure, 207, 275, 279, 283, 286
- set intersection, 126
  - binary search, 75

- compressed storage, 80
- doubling search, 77
- lower bound, 77
- merge-based, 74
- mutual partitioning, 75
- optimal time bound, 77, 78
- two-level storage, 79
- Shannon's coding theory, 196
- Shannon, Claude, 196
- simulation, parallel algorithm, 35
- skew algorithm, 162
- sorting, 35
  - atomic items, 44
  - binary MERGESORT, 45
  - binary QUICKSORT, 57
  - ColumnSort, 69
  - CountingSort, 87
  - distinguishing (string) prefix, 83
  - HEAPSORT, 47
  - INSERTIONSORT, 47
  - lower bound, 52, 55, 83
  - multi-disks, 67
  - multi-key QuickSort, 90
  - multi-way MERGESORT, 50
  - multi-way QUICKSORT, 65
  - RADIXSORT, 84
  - stable, 87
  - string suffixes, 162, 168, 266
  - strings, 82, 162
- statistical coding
  - arithmetic, 221
  - asymmetric numeral systems (ANS), 229
  - Huffman, 210
  - Huffman, canonical, 217
  - prediction by partial matching (PPM), 235
  - range coding, 230
- String B-tree, 146
- string prefix search
  - array of string pointers, 129
  - Patricia trie, 142
  - String B-tree, 146
  - ternary search tree, 93
  - trie, (un)compacted, 138
  - two-level indexing, 133, 141, 146
- string substring search, 153
  - algorithmic reduction, 154
  - longest common prefix (LCP), 158
  - suffix, 153
  - suffix array, 154
  - suffix tree, 175
- string, variable-length key, 82, 153
- subarray sum, maximal, 10
- suffix array, 154, 176, 254
- suffix tree, 172, 190
- technique
  - Burrows–Wheeler transform, 252
  - coin tossing, 40, 41
  - compressed graph encoding, 291
  - compression booster, 267
  - disk striping, 67
  - divide-and-conquer, 38, 45, 57, 75, 84, 90, 162
  - four Russians trick, 188
  - global rebuilding, 100, 116
  - LZ text parsing, 189, 240
  - Move-to-Front Transform, 259
  - multi-way MergeSort, 50
  - pivot selection, 65
  - pointer jumping, 33
  - prediction by partial matching (PPM), 235
  - Run Length Encoding transform, 263
  - Snow Plow, 48
  - three-way partitioning, 59
  - tree encoding, succinct, 284
  - tree packing on disk, 146
- ternary search tree, 93
- text mining, 191
- tree
  - canonical, Huffman, 217
  - disk packing, 146, 149
  - Huffman, 212
  - $k^2$ -ary tree, 295
  - succinct representation, 284
  - suffix tree, 172
  - ternary search tree, 93
- trie, 84, 138, 245
  - compacted, 141
  - disk packing, 146
  - Huffman, 212
  - Patricia trie, 142
  - String B-tree, 146
  - uncompacted, 139, 245
- unary integer code, 196
- variable-byte integer code, 200
- Webgraph compressor, 292
- Wheeler, David, 252
- Ziv, Jacob, 189, 240