

# *Index*

acceptor site, 200  
additive matrix, 358  
additive tree, *see* additive matrix  
adjacency, 132  
affine gap penalty, 184  
Aho-Corasick algorithm, 328  
algorithm, 7, 17  
    ADDITIVEPHYLOGENY, 362  
    ALLLEAVES, 103  
    APPROXIMATEPATTERNMATCHING,  
        326  
    approximation, 21, 131  
    approximation ratio, 131  
    Baum-Welch, 396  
    BETTERCHANGE, 21  
    BRANCHANDBOUNDMEDIANSEARCH,  
        114  
    BRANCHANDBOUNDMOTIFSEARCH,  
        111  
    BREAKPOINTREVERSALSORT, 133  
    BRUTEFORCECHANGE, 22  
    BRUTEFORCEMEDIANSEARCH, 112  
    BRUTEFORCEMOTIFSEARCH, 109  
    BRUTEFORCEPDP, 88  
    BYPASS, 108  
    CAST, 352  
    clustering, 338  
    correctness, 17, 20, 21  
    DPCHANGE, 151  
    DUPLICATEREMOVAL, 311  
    efficiency, 33, 151  
    EXONCHAINING, 202  
    FASTROCKS, 47  
    FIBONACCI, 33  
    Fitch's, 370  
    GREEDYMOTIFSEARCH, 136  
    GREEDYPROFILEMOTIFSEARCH, 410  
    HANOITOWERS, 26  
    HIERARCHICALCLUSTERING, 343  
    IMPROVEDBREAKPOINTREVERSAL-  
        SORT, 135  
    incorrect, 21  
        BETTERCHANGE, 21  
        BREAKPOINTREVERSALSORT, 133  
    LCS, 176  
    MANHATTANTOURIST, 160  
    MERGE, 228  
    MERGESORT, 228  
    NEIGHBORJOINING, 365  
    NEXTLEAF, 103  
    NEXTVERTEX, 107  
    origins of, 51  
    PARTIALDIGEST, 90  
    PATH, 234  
    PATTERNMATCHING, 315  
    PCC, 350  
    PRINTLCS, 176  
    PROGRESSIVEGREEDYK-MEANS, 346  
    QUICKSORT, 408  
    RANDOMIZEDQUICKSORT, 409  
    RANDOMPROJECTIONS, 414  
    RECURSIVECHANGE, 150

- RECURSIVEFIBONACCI, 33
- RECURSIVESELECTIONSORT, 30
- ROCKS, 47
- Sankoff's, 369–370
- SELECTIONSORT, 29
- SIMPLEREVERALSORT, 129
- STRINGCOPY, 16
- SUFFIXTREEPATTERNMATCHING, 321
- UPGMA, 364
- USCHANGE, 19
- alignment, 168
- Alu repeat, 276
- amino acid, 65
- anticodon, 67
- apoptosis, 283
- ATP binding proteins, 148
- Avery, Oswald, 61
  
- BAC, 277
- backtracking pointers, 175
- backward algorithm, 395
- bacteriophage, 260
- balanced partitioning, 244
- balanced vertex, 273
- Beadle, George, 60
- Benzer, Seymour, 260
- BLAST
  - maximal segment pair, 328
  - statistical significance, 328
- block alignment, 234–237
- BLOSUM matrix, 178
- Bohr, Niels, 5
- border length of mask, 304
- Braconnot, Henry, 65
- branch-and-bound algorithms, 42
- breakpoint, 132
  
- cells, 57
  - eukaryotic, 63
  - prokaryotic, 63
- central dogma, 67
- CG-island, 385
  
- Chargaff rule, 62
- Chargaff, Erwin, 61
- chromosome, 59
- cloning, 68
- cloning vector, 68
- CLUSTAL, 191
- codon, 65
  - Start, 197
  - Stop, 197
- complement, *see* Watson-Crick complement
- computer language, 7
- CONSENSUS, 83, 136
- consensus (in fragment assembly), 279
- consensus string, 93
- Crick, Francis, 5, 6, 61
- cystic fibrosis, 148
  
- DAG, *see* graph, DAG
- David Haussler, 401
- David Sankoff, 139
- DDP, *see* problems, Double Digest Problem
- degree, *see* vertex, degree
- Delbrück, Max, 5
- deletion, 169
- divide-and-conquer algorithms, 48
- DNA, 58
  - array, 265, 337
  - helicase, 14
  - nucleotides, 61
  - polymerase, 15
  - read, 275
  - reads, 263
  - repeat, 276
  - replication of, 14–16
    - discovery, 63
    - ligase, 15
    - Okazaki fragments, 15
    - primers and primase, 15
    - replication fork, 14
    - SSB, 14

- vector, 263
- donor site, 200
- dot matrix, 323
- double helix, 61
- Drosophila melanogaster*, 59
- dynamic programming, 43, 176
- edge, 153, 248
- edit distance, 167, 207
- edit graph, 170
- Edman degradation, 280
- EDVAC, 240
- emission probability, 388
- entropy score, 193
- Eulerian cycle, 258, 272–275
- Eulerian graph, 273
- evolution, 74
- evolutionary character, 366
- exhaustive search, 41
- exon, 63, 194
- experimental spectrum, 285
- exponential algorithm, 35
- FASTA, 323
- Fibonacci numbers, 31
- filtration
  - database search, 323, 327
  - of candidate exons, 207
- Fischer, Emil Hermann, 65
- fitting alignment, 215
- floor function, 20
- forward algorithm, 394
- Four Russians technique, 237
- four-point condition, 363
- gap, 184
  - once a gap, always a gap, 191
- gap penalty, 184
- Gary Stormo, 116
- Gates, William, 131
- gel electrophoresis, 72, 85
- gene, 60
- Gene Myers, 331
- generating function, 119
- genes
  - immunity, 91
- genetic code, 65
- genetic linkage, 60
- genome rearrangement, 126
  - reversal, 127, 128
  - reversal distance, 128–129
- GENSCAN, 200
- giant panda, 352
- Gibbs sampling, 410
- GibbsSampler, 83
- Gilbert, Walter, 263
- global alignment, 177
- graph, 153, 248
  - clique, 346
  - complete, 252
  - connected, 252
  - connected component, 252
  - DAG (directed acyclic graph), 162
  - directed, 250
  - disconnected, 252
  - edge, *see* edge
  - interval, 262
  - path in a, *see* path
  - vertex, *see* vertex
- greedy algorithms, 43
- Hamiltonian cycle, 258, 272
- Hamming distance, 98, 167
  - total, 98
- hexamer count, 199
- Hidden Markov Model, *see* HMM
- hidden state, 388
- hierarchical clustering, 342
- HMM, 388
  - parameter estimation, 395
  - path, 389
  - profile HMM, 396–398
- homeobox genes, 180
- homometric sets, 87
- Hooke, Robert, 57

- Human Diversity Project, 73
- Hurwitz, Jerard, 64
- hybridization, 72, 267
- indel, 169
- insert (DNA), 263
- insertion, 169
- insulin, 280
- intron, 63, 195
- ion type, 285
  - b-ion, 285
  - y-ion, 285
- junk DNA, 194
- $k$ -similarity, 295
- $l$ -mer composition, 267
- layout, 279
- layout of DNA fragments, 279
- Leonhard Euler, 299
- Levenshtein distance, *see* edit distance
- light-directed array synthesis, 303
- likelihood ratio test, 198
- LINE repeat, 277
- Linear B, 75
- local alignment, 181
- machine learning, 48
- mask for array synthesis, 303
- mass spectrometry, 283
  - description, 283
  - MS/MS, *see* mass spectrometry, tandem
  - tandem, 284
- mass spectrum, 283
- match, 169
- mates, 279
- maximization algorithm, 131
- Mendel, Gregor, 59
- Michael Waterman, 209
- Miescher, Johann Friedrich, 61
- minimization algorithm, 131
- mismatch, 169
- Morgan, Thomas, 59
- mosaic effect, 207
- multiple alignment, 188
- multiset, 84
- mutation, 184
- mutations, 93
- NF- $\kappa$ B binding motif, 92
- nonadditive matrix, 358
- $\nu$ -sis oncogene, 147
- $O$ -notation, 37
- Occam's razor, 366
- open reading frame, 197
- optical mapping, 307
- ORF, *see* open reading frame
- Out of Africa hypothesis, 353
- overlap, 279
- PAM matrix, 178
- Papadimitriou, Christos, 131
- path, 153, 252
  - cycle, 252
- path length, 153
- pattern-driven approach, 93
- PCR, 67
- PCR primer, 68
- peptide
  - C-terminal, 284
  - N-terminal, 284
- performance guarantee, 131
- permutation, 127
- Poe, Edgar Allan, 92
- polynomial algorithm, 35
- post-translational modification, 291, 338
  - glycosylation, 291
  - phosphorylation, 291
- probe, 73
- problem, 7
  - complexity, 151
  - input, 7
  - instance, 17

- output, 7
- problems
  - $k$ -Means Clustering Problem, 344
  - Approximate Pattern Matching Problem, 325
  - Block Alignment Problem, 237
  - Bridge Obsession Problem, 254
  - Change Problem, 21
  - Chimeric Alignment Problem, 219
  - Corrupted Cliques Problem, 348
  - Decoding Problem, 391
  - Distance-Based Phylogeny Problem, 358
  - Double Digest Problem, 121
  - Duplicate Removal Problem, 311
  - Equivalent Words Problem, 211
  - Eulerian Cycle Problem, 254
  - Exon Chaining Problem, 201
  - Fair Bet Casino Problem, 386
  - Fibonacci Problem, 31
  - Global Alignment Problem, 177
  - Hamiltonian Cycle Problem, 258
  - Inexact Repeat Problem, 218
  - Knights Tour problem, 250
  - Large Parsimony Problem, 372
  - Least Squares Distance-Based Phylogeny Problem, 364
  - Lewis Carroll Problem, 211
  - Local Alignment Problem, 181
  - Longest Common Subsequence Problem, 173
  - Longest Path in a DAG Problem, 162, 288, 289
  - Manhattan Tourist Problem, 154
  - Median String Problem, 99
  - Modified Protein Identification Problem, 291
  - Motif Finding Problem, 98
  - Multiple Breakpoint Distance Problem, 145
  - Multiple Pattern Matching Problem, 316
  - $\mathcal{NP}$ -complete, 49, 348, 372
  - Pancake Flipping Problem, 130
  - Partial Digest Problem, 86
  - Pattern Matching Problem, 315
  - Peptide Sequencing Problem, 287
  - Probed Partial Digest Problem, 121
  - Protein Identification Problem, 290
  - Query Matching Problem, 326
  - Reversal Distance Problem, 128
  - Sequencing by Hybridization (SBH) Problem, 270
  - Shortest Path Problem, 258
  - Shortest Superstring Problem, 264
  - Small Parsimony Problem, 368
  - Sorting by Reversals Problem, 129
  - Sorting Problem, 29
  - Spectral Alignment Problem, 295
  - Spliced Alignment Problem, 204
  - String Duplication Problem, 16
  - Tandem Repeat Problem, 218
  - Towers of Hanoi Problem, 24
  - Traveling Salesman Problem, 50, 258, 264
  - Turnpike Problem, 86
  - United States Change Problem, 19
  - Weighted Small Parsimony Problem, 368
- profile, 396
- profile HMM alignment, 396
- protein, 58
- protein complex, 66
- protein sequencing, 300
- pseudocode, 8
  - argument, 8
  - array, 8
  - operations, 8–11
  - subroutine, 8, 13
  - variable, 8, 13
- pumpkin pie, 11
- quadratic algorithm, 35
- randomized algorithms, 48

- RandomProjections, 83
- rearrangement scenario, 127
- red panda, 352
- restriction enzyme, 71, 83
- restriction fragment, 83
- restriction fragments, 71
- restriction map, 83
- ribosome, 66
- Richard Karp, 52
- RNA, 58
  - folding, 221
  - messenger RNA (mRNA), 64
- Rocks game, 1
- Ron Shamir, 378
- Rosetta stone, 77
- Russell Doolittle, 79
  
- Sanger, Frederick, 263, 280
- SBH, *see* Sequencing by Hybridization
- Schleiden, Matthias, 57
- Schrödinger, Erwin, 5
- Schwann, Theodor, 57
- score of multiple alignment, 188
- scoring matrix, 153, 178–180
- semibalanced graph, 275
- Sequencing by Hybridization, 265
- SEQUEST, 290
- shared peaks count, 285
- shotgun sequencing, 263
- simian sarcoma virus, 147
- similarity score, 176
- Smith, Hamilton, 83
- Smith, Temple, 180
- SP-score (Sum-of-Pairs), 193
- spectral alignment, 293
- spectral convolution, 292
- spectral product, 296
- spectrum graph, 287
- spliced alignment graph, 205
- splicing, 64, 195
  - alternative, 66
- state transition probability, 388
  
- statistical distance, 335
- strip, 133
  - decreasing, 133
  - increasing, 133
- Sturtevant, Alfred, 60
- subsequence, 172
- supersequence, 217
- superstring, 264
- synteny blocks, 126
  
- Tatum, Edward, 60
- test data, 385
- theoretical spectrum, 285
- tiling arrays, 271n
- topological ordering, 164
- Towers of Hanoi, 24
- training data, 385
- translation, 67
- tree, *see* DAG
- tree traversal
  - ALLEAVES, 103
  - BYPASS, 108
  - in-order, 370
  - NEXTLEAF, 103
  - NEXTVERTEX, 107
  - post-order, 370
  - pre-order, 370
  - pruning, 107
- triplet rule, 65
  
- universal DNA array, 267
  
- vertex, 153, 248
  - degree, 250
  - indegree, 162, 250
  - outdegree, 162, 252
  - predecessor, 163
- von Neumann, John, 240
  
- Waardenburg's syndrome, 125
- Waterman, Michael, 180
- Watson, James, 5, 6, 61
- Watson-Crick complement, 266

Webb Miller, 241  
weighted interval, 201  
Weiss, Samuel, 64  
worst case efficiency, 37  
  
Zamecnik, Paul, 63

