

General Index

- acceptance window, 563
- acoustic model, 231
- adjacency list, 109
- agglutinative language, 176
- algebraic power series, 537
- algorithm
 - ACYCLICMINIMIZATION, 34
 - ADDTOTRIE, 17
 - AUTOMATAPRODUCT, 39
 - AUTOMATAUNION, 38
 - AUTOMATONLETTER, 38
 - AUTOMATONSTAR, 39
 - BORDER, 8
 - BORDERSHARP, 97
 - CHECKSIMPLE, 499
 - CIRCULARMIN, 14
 - CLONE, 133
 - CLOSURE, 510
 - COMPOSETRANSDUCERS, 44
 - CONVEX, 500
 - DOMINANTEIGENVALUE, 73
 - ENCODEMAP, 516
 - ENTROPY, 90
 - EPSILON, 58
 - EVALEXP, 56
 - EVALFACT, 56
 - EVALTERM, 56
 - EXTENSION, 133
 - FASTFIND, 118
 - FIRST, 59
 - FLORENTINEREJECTION, 503
 - FOLLOW, 60
 - FOOTPRINTER, 253
 - FORBIDDENWORDS, 152
 - HOPCROFTMINIMIZATION, 31
 - ISACCEPTED, 23
 - ISINTRIE, 17
 - ISSUBWORD, 11
 - K-REPETITIONS, 468
 - LCP, 48
 - LCS, 13
 - LCSENGTHARRAY, 12
 - LENGTHSOFFACTORS, 156
 - LLTABLE, 62
 - LONGEST-PREFIX-EXTENSION, 441
 - LONGESTCOMMONPREFIX, 7
 - LONGESTCOMMON-PREFIXARRAY, 51
 - LONGESTPREFIXINTRIE, 17
 - LRPARSE, 67
 - LYNDONFACTORIZATION, 15
 - MAXIMAL-REPETITIONS, 447
 - MISMATCH-RIGHT-REPETITIONS, 465
 - MISMATCH-RIGHT-SUBRUNS, 469
 - MOOREMINIMIZATION, 28
 - NAIVESTRINGMATCHING, 10
 - NEXT, 21–23, 48
 - NFATODFA, 24
 - NORMALIZETRANSDUCER, 51
 - PROCESS-BLOCK, 467
 - RANDMAP, 511
 - RANDPERM, 494
 - REMOVEFROMTRIE, 18
 - RIGHT-LOCAL-SQUARES, 459
 - RIGHT-REPETITIONS, 444

- algorithm (*cont.*)
 RUNS-THIRD-STAGE, 472
 SEARCHFACTOR, 11
 SLOWFIND, 109
 SLOWFIND-BIS, 110
 SLOWFINDC, 117
 SMILE, 258
 SPELLER, 251
 STAIRCASE, 496
 SUFFIXAUTOMATON, 132
 SUFFIXTREE, 116
 SUFFIXTRIE, 109
 SUFFIXTRIE-BIS, 111
 TOSEQUENTIALTRANSUCER,
 49
 UNIMODAL, 498
 VITERBI, 98
 WEIGHTED-COMPOSITION,
 213
 WEIGHTED-
 DETERMINIZATION, 218
 ZLDECODING, 87
 ZLENCODING, 86
- alignment, 188, 189
 alphabet, 3
 DNA alphabet, 270
 alphabetic order, 5
 ambiguity
 lexical, 173
 amino acid, 271
 animal, 492
 approximate eigenvector, 73
 approximate repetition, 461
 approximate square, 461
 Arnoux–Rauzy word, 575
 asymptotic equirepartition property,
 85
 autocorrelation
 polynomial, 93, 278, 311, 361,
 402, 415
 set, 361
 automatic sequence, 524
 automaton, 19
 complete, 21
 deterministic, 20
 finite, 21
 literal, 20
 minimal, 27
 of analyses, 198
 path, 19
 state, 19
 trim, 19
 unambiguous, 19
 weighted, 212
- basis, 260, 261, 267
 beam search, 234
 synchronous, 234
 Bernoulli distribution, 75
 Bernoulli source, 356
 best approximations, 564
 beta
 expansion, 572, 577
 integers, 573
 transformation, 572, 577
 bimachine, 103, 192
 bounded remainder
 letter, 570
 set, 571
- capacity
 (l, k) sequences, 393
 Carlitz formal power series, 541
 cascade transduction, 186
 Catalan number, 70, 484
 central limit theorem, *see* Gaussian
 approximation
 exact string matching, 370
 generalized pattern matching, 389
 reduced string matching, 379
 self-similar pattern matching,
 415
 subsequence pattern matching,
 402
- cepstra, 232
 delta, 232
 delta–delta, 232
 cepstral coefficients, 232
 cepstrum, 232
 character
 encoding, 165
 Chen–Stein method, 339

- chi-square
 - statistic, 276, 277
 - test, 275
- Christol's theorem, 539
- clump, 278, 279
 - k -clump, 278, 280
 - mixed, 322
 - size of, 278
- codon, 271
- COMBI, 244, 252, 260, 262, 265
- compact suffix automaton, 138
- compaction of trie, 113
- complexity function, 569
- composition
 - filter, 215
 - of weighted transducers, 213, 234
- compound Poisson approximation, 302, 319, 340
- compound words, 172
- confidence interval, 307, 340
- conjugate words, 4, 159
- constant length morphism, 522
- constituent, 204
- context-dependency
 - model, 230
 - n -phonic model, 230
 - transduction, 230
- context-free grammar, 204, 206
- correlation
 - matrix, 377
 - number, 398
 - polynomial, 324, 377
 - set, 377
- count
 - of clumps, *see* declumped count
 - of overlapping occurrences, *see* word count
 - of renewals, *see* renewal count
- critical factorization theorem, 455
- cut and project scheme, 563
- DAWG, 174
- de Bruijn graph, 229, 335, 382
- declumped count, 279, 298
 - Poisson approximation, 298, 302, 320
- decoder, 234
- delimiter, 168
- delta method, 294, 344
- density, 534
- depth, 416
- deterministic
 - weighted automata, 217
- determinizable
 - weighted automata, 220
- determinization, 181
 - weighted automata, 217
 - weighted transducers, 217, 235
- dictionary, 172
- discrete topology, 521
- distance, 5
 - edit, 6, 243, 262, 264, 266
 - Hamming, 5, 243, 245, 266
 - subword, 6
- distance between occurrences, 282
 - distribution, 283
 - expectation, 284
 - moment-generating function, 284
 - scan, 286
 - variance, 284
- don't care symbol, 241, 242, 244, 259–261
- Dyck language, 53
- dynamical source, 357, 407
- edge label representation, 114
- encoding
 - character, 165
- end position, 122
- error
 - lexical, 176
- exact pattern matching, 359
- exceptional word, 328
- exponent, 431
- factor, 3
 - proper, 3
- failure function, 155, 158
- family of words, 315, 319, 322, 323
 - reduced set, 316
- Fibonacci numbers, 21, 524
- Fibonacci word, 524

- Fibonacci words, 434
- Fine and Wilf's theorem, 432
- fixed point, 522, 524
- FOOTPRINTER, 245, 252, 253, 264, 266
- forbidden word, 152
- fork, 108, 138
- formal Laurent series, 536
- formal power series, 536
- Fourier transform, 232

- Gamma function $\Gamma(z)$, 367
- Gaussian approximation, 290, 312, 316, 324, 338
- generalized index, 143
- generalized pattern matching, 375
- generating operator, 359, 407
- Gilbreath's card trick, 25
- Golden mean automaton, 21, 570
- grammar
 - statistical, 227
 - weighted, 227
- graph, 179, 201

- hapax, 170
- head, 107
- hidden Markov model, 231, 235
- hidden patterns, 355, 393
- homography, 173
- Hopcroft's algorithm, 30

- implementation of transition, 109
- incidence matrix, 535, 546
- index, 142
- index membership, 144
- inflection, 172
- initial prefix, 45
- integer power, 431
- intersection
 - weighted automata, 215
- irreducibility, 384
- iterates of a morphism, 522

- kernel, 526, 527
- KMRC, 246, 266
- Kolmogorov theorem, 78

- lacuna
 - lexical, 176
- language, 362
- language model, 227
 - Katz back-off, 229, 235
- large deviation principle, 309, 345
- large deviations
 - exact string matching, 371
 - generalized pattern matching, 389
 - reduced string matching, 379
 - subsequence pattern matching, 405
- lattice, 198
- lemma, 172
- lexical
 - ambiguity, 173
 - analysis, 165
 - error, 176
 - lacuna, 176
 - tagging, 172
- lexicographic order, 5
- ligature, 165, 168
- likelihood, 273
- literal
 - alignment, 188
- literal automaton, 20
- local grammar, 201
- local limit theorem
 - exact string matching, 370
 - generalized pattern matching, 391
 - subsequence pattern matching, 405
- local period, 455
- longest common
 - factor, 159
 - prefix, 3
 - subword, 3
- longest context, 145
- longest extension functions, 439
- loop, 480

- Markov chain, 75, 198, 269, 271
 - aperiodic, 76
 - embedding technique, 273, 308
 - irreducible, 76
 - maximal order, 291

- order, 271, 275, 329
- parameters, *see* transition matrix
- phase, 272
- reversed chain, 273
- stationary distribution, 271
- Markov source, 357, 366
- martingale, 291, 345
- matrix
 - aperiodic, 71
 - irreducible, 71
 - nonnegative, 71
 - positive, 71, 535
 - primitive, 71
 - stochastic, 535
- maximal repetition, 432
- memoryless source, 356
 - biased, 356
 - unbiased, 356
- Meyer set, 563
- minimal
 - weighted automata, 223
- minimal automaton, 27
- minimal forbidden word, 152
- minimization
 - weighted automata, 223
 - weighted transducers, 235
- MITRA-COUNT, 244, 252, 255, 264–266
- MITRA-DYAD, 257, 267
- MITRA-GRAPH, 246, 255–257, 264
- model, *see* probabilistic model
- model set, 563
- Möbius, 567
- moment-generating function, 284, 343
- moments
 - exact string matching, 366
 - generalized subsequence problem, 411
 - subsequence pattern matching, 400
- Moore's algorithm, 28
- morphic, 524
- morphism, 522
 - of constant length, 522
 - of Pisot type, 570, 576
- uniform, 522, 524
- unimodular, 576
- network expression, 241–249, 251–265, 267
 - basis, 260, 261, 267
 - simple, 241–245, 247–249, 251–256
 - tandem, 261–264
 - with spacers, 241, 256–260
- NEWAUTOMATON, 109
- NEWSTATE, 109
- normal approximation, *see* Gaussian approximation
- number of factors, 147
- occurrence, 277, 278
 - distance, *see* distance between occurrences
 - overlap, *see* period
 - phased, 296
 - probability, 279
 - waiting time, 347
- output of state, 109
- overlap, 546
- overlapping occurrences, 277
 - clump, *see* clump
- paperfolding word, 526, 528
- path, 19, 492
 - successful, 19
- period, 4, 277
 - minimal, 4
 - principal, 277, 279
- Perron–Frobenius theorem, 70, 380, 386, 535
- perturbed symmetry, 526
- phase, 296, 332
- phone, 227, 230
 - context-dependent, 230
- phoneme, 230
- Pisot number, 546
- Poisson approximation, 287, 297, 313, 319, 325, 336, 339
- POIVRE, 244, 247, 252, 260, 262, 264–266

General Index

609

- solid, 130
- source
 - memoryless, 356
- spacer, 241, 256, 257, 259, 260, 265, 267
- spectral decomposition, 388
- spectrum, 335
- speech, 196, 226
- speech recognition, 226
 - statistical formulation, 226
- SPELLER, 244, 249, 251, 252, 255, 257–260, 263–266
- square, 431
- square-free number, 568
- state, 19
 - initial, 19
 - terminal, 19
- state diagram, 19
- statistical
 - grammar, 227
- statistical grammar
 - Katz back-off, 229
 - n -gram model, 228
- Stein's method, 338, 340
- step, 480
- stochastic matrix, 535
- string matching
 - exact, 354, 359
 - generalized, 354, 375, 382
 - reduced, 376
 - self-similar, 413
- structured motif, 347
- Sturmian word, 524
- subsequence pattern matching, 355, 393
 - generalize, 355
 - generalized, 407
- subsequentiability
 - weighted automata, 219
- subsequential
 - weighted automata, 217
- substitution matrix, 535
- substitutive, 524
- subword, 3
- suffix, 3
 - proper, 3
- suffix array, 163
- suffix automaton, 121, 127
- suffix function, 123
- suffix link, 110, 130
- suffix link function, 110
- suffix paths, 131
- suffix tree, 113, 248, 249, 251, 259, 266
- suffix trie, 107, 416
- syntactic
 - table, 205
- syntactic congruence, 122
- syntactic equivalence, 121
- table
 - syntactic, 205
- tag, 171
- tail, 108
- TARGETSTATE, 109
- terminal
 - function, 45
- Thue–Morse word, 77, 525, 528
- token, 168
- topological entropy, 82
- topology, 521
- total variation distance, 339
- transducer, 41
 - p -sequential, 173
 - alignment, 188
 - ambiguous, 173, 180
 - finite, 166
 - generalized sequential, 181
 - literal, 42, 166
 - sequential, 45, 167
 - subsequential, 103
 - synchronous, 42, 166
 - weighted, 212
- transduction, 166
 - cascade, 186
- transition matrix, 271, 535
 - eigenvalue, 272
 - estimation, 273
- transliteration
 - ambiguous, 168
 - unambiguous, 166

- Tribonacci
 - morphism, 546
 - normal representation, 548
 - sequence, 548
 - word, 525, 546
- trie, 416, 423
- twinning property, 47
- twins
 - property, 220
 - states, 220
- two-level morphology, 191

- uniform balance, 570
- uniform morphism, 522, 524
- unique ergodicity, 87

- vector
 - nonnegative, 70
 - positive, 70
- Viterbi
 - algorithm, 98, 234
 - approximation, 211, 233, 234

- waiting time, 359, 374
- walk, 480
 - reverse, 480
 - length of a, 480
 - self-avoiding, 492
- WEEDER, 246, 266
- weight pushing, 220, 235
 - weighted automata, 220
- weighted automata
 - deterministic, 217
 - determinizable, 220
 - intersection, 215
 - minimal, 223
 - minimization, 223
 - subsequential, 219
 - subsequential, 217
 - weight pushing, 220
- weighted transducers
 - composition, 213, 234
 - determinization, 217
 - minimization, 223
- Wielandt function, 72
- WINNOWER, 246, 247, 254, 255, 264, 266
- word
 - conjugate, 4
 - Dyck, 482
 - linguistic, 172
 - Motzkin, 485
 - palindrome, 4
 - period, 4
 - primitive, 4
 - reversal, 4
- word bag model, 199
- word count, 279
 - asymptotic variance, 294
 - compound Poisson
 - approximation, 302, 306, 307, 322
 - distribution, 289
 - expectation, 290
 - Gaussian approximation, 290, 316
 - variance, 290
- Z-score, 373
- Ziv–Lempel encoding, 86
- Ziv–Lempel factorization, 86