

Index

- p -norms, 101, 103
 - ∞ -norm (Max-norm, Chebyshev distance), 101, 104
 - L1 (Manhattan, city block), 104, 209, 239, 265, 303, 305, 329, 369
 - L2 (Euclidean), 103, 105, 163, 238, 239, 265, 272, 303, 305, 329
- Approximate sequence matching, 195
 - Deterministic finite automaton, 195
 - Filtering, *see* String fingerprinting, 198
 - Multiple, *see* Multiple approximate sequence matching, 206
 - Non-deterministic finite automaton, 195, 206
- Aspect model
 - Latent semantics, 176
 - Log-likelihood function
 - Maximization, 176
- Audio models, 90
 - Beat/tempo, 90
 - Beat/tempo
 - Beats per minute (BPM), 91
 - Periodicity, 91
 - Chroma, 90
 - Pitch perception, 91
 - Frequency, 164
 - Loudness, 90
 - Phon, 91
 - Signal power, 91
 - Sone, 91
 - Onset strength signal, 91
 - Pitch, 90, 164
 - Bark scale, 91
 - Frequency analysis, 90
 - Frequency analysis, *see* Signal, 90
 - Mel (melody) scale, 91
 - Rhythm, 90, 91
 - Rhythmogram, 92
 - Timbre, 90, 92
 - Timbre
 - Timbregram, 92
 - Volume, 164
- Bayesian models
 - Bayesian inference, 136
 - Bayesian information criterion (BIC), 323
 - Bayesian networks, 123, 136
 - Bayesian networks
 - Hidden Markov models (HMM), *see* Hidden Markov models (HMM), 136
 - Bayesian theorem, 136, 314
 - Bayes' rule, 135
 - Classification, 314
 - Expectation maximization (EM), *see* Expectation maximization (EM), 136
 - Generative models, *see* Generative models, 137
 - Independence assumption, 314
 - Relaxing, 315
 - Language models, *see* Language models, 136
 - Markov Chain Monte Carlo (MCMC), 136
 - Maximum likelihood estimation (MLE), *see* Maximum likelihood estimation (MLE), 314
 - Query answering, 136
- Classification, 297
 - Bayesian, *see* Bayesian models, 314
 - Boosting, 324
 - Boosting
 - AdaBoost, 324, 421
 - Class label, 297
 - Decision trees, 297, 309
 - Fisher's discriminant ratio, *see* Fisher's discriminant ratio, 299
 - Fuzzy, 311
 - Information gain, *see* Information gain, 298
 - Random forest, 300
 - Ensemble, 300, 324
 - Fuzzy, 311, 312
 - Hidden Markov models (HMM), 316, *see* Hidden Markov models (HMM), 316
 - Linear discriminant analysis, 301
 - Least absolute regression deviation (LAD), 302

474 Index

- Classification (*cont.*)
 - Max-margin learning, *see* Support vector machines, 303, 308
 - Nearest neighbor classifiers, 301
 - Overfitting, 300–302, 322
 - Model selection, 322, *see* Model selection, 322
 - Random forest, 300
 - Bagging, 300
 - Random split selection, 300
 - Random subspace, 300
 - Regularization, 302
 - L1-, 303
 - L2-, 303
 - Least absolute shrinkage and selection operator (LASSO), 303
 - Ridge-regression, 303
 - Rule-based, 308
 - Assertions, 309
 - Decision trees, 309
 - Fuzzy, 311, 312
 - Pruning, 310
 - Rules, 309
 - Simplification, 310
 - Strong classifier, 324
 - Supervised learning, 297
 - Support vector machines, *see* Support vector machines, 301
 - Training set, 297
 - Unsupervised learning, 297
 - Voted perceptron, 308, 324
 - Batch, 308
 - On-line, 308
 - Voting-based, 301, 308
 - Weak classifier, 324
- Clustering, 105, 271
 - Agglomerative, 282
 - Agglomerative
 - Bottom-up, 282
 - Hierarchical, 282
 - Cluster centroid, 273, 283
 - Cluster diameter, 272
 - Co-clustering, *see* Co-clustering, 292
 - Compactness, 272
 - Dynamic evidence, 288
 - Estimating the number of clusters, 284
 - Covering, 284
 - G-means, 286
 - Incremental, 285
 - Probabilistic, 284
 - X-means, 285
 - Graph-based, 275
 - Adaptive thresholding, 279
 - Angular clustering, 277
 - Connected components, 275
 - Maximal cliques, 276
 - Minimum cut-based, 46, 278
 - Random walk-based, 142, 277
 - Spectral partitioning, 277
 - Iterative methods, 280
 - Multipass, 282
 - Single pass, 280
 - K-means, 46, 270, 282
 - FM algorithm, 283
 - G-means, 286
 - KL algorithm, 283
 - X-means, 285
 - Mixture Models, 287
 - Multipass
 - Agglomerative, 282
 - Iterative improvement, 282
 - K-means, 282
 - Multi-constraint, 286
 - Horizontal formulation, 287
 - Vertical formulation, 287
 - Online, 288
 - Confidence-based, 288
 - Perturbation-based, 289
 - User access patterns, 288
 - Quality measures, 272, 283
 - Cheeger's ratio, 273
 - Cluster compactness, 272
 - Cluster homogeneity, 272
 - Cluster integrity, 274, 279
 - Cluster overlap, 273
 - Cluster separation, 273, 279
 - Cluster size, 273, 281
 - Clustering balance, 273
 - Clustering modularity, 274
 - Conductance, 274
 - Cut-based, 273
 - Entropy, 273, 281, 296
 - Expansion-based, 274
 - Root-mean-square-error (RMSE), 272
 - Sum-of-squares, 272
 - Self-organizing maps (SOM), 290
 - Cortex, 290
 - Distance preserving, 292
 - Neural networks, *see* Neural networks, 290
 - Single pass, 280
 - Adaptive leader selection, 281
 - Cluster quality, 281
 - Leader, 281
 - Max-a-min leader selection, 281
 - Scan order, 281
 - Threshold value, 282
 - Co-clustering, 292
 - Column clustering, 294
 - Cross-association, 294
 - Information theoretical, 293
 - LSI, 292
 - Row clustering, 294
 - Collaborative filtering, 1, 8, 10, 229, 230, 400, 413
 - AdaBoost, 325, 421
 - Associative retrieval-based, 417
 - Classification-based, 229, 414, 418
 - Clustering-based, 418
 - Data sparsity problem, 416
 - Graph model, 413
 - Horting, 416
 - Item-based, 414
 - Memory-based, 414, 415, 419
 - Model-based, 414, 418, 419
 - Nearest neighbor-based, 416
 - Object metadata, 229, 413
 - Preferences, 229
 - RankBoost, 423
 - Ratings, 229
 - Combining, 421
 - Spreading activation, 417

- Top-K retrieval-based, 414
- User metadata, 229, 413
- User-based, 414
- Voting matrix, 413
- Voting-Based, 415
- Color models, 99
 - Brightness, 38
 - Chroma, 38, 39
 - Energy, 39
 - Saturation, 39
 - CIELAB model
 - CIE space, 37
 - Lab space, 38
 - Color channels (components)
 - Blue (B), 35
 - Chrominance (I,Q), 37
 - Chrominance (U,V), 37
 - Green (G), 35
 - Hue (H), 39
 - L, a, b, 38
 - Luminance (Y), 37
 - Red (R), 35
 - Saturation (S), 39
 - Value (V), 39
 - X,Y,Z, 38
 - Color instance
 - Bit representation, 35
 - Color similarity, 107
 - Color table, 35
 - Median-cut algorithm, 36
 - Contrast, 32, 37
 - Distinct colors, 35
 - Histogram, 40, 107, 149
 - Histogram, *see* Histogram, 40
 - Matching, 40
 - HSV space, 39
 - RGB space, 35, 149
 - Color cube, 35
 - YIQ space, 37
 - YRB space, 37
 - YUV space, 37
- Common factor analysis (CFA), 160, 161
 - Error terms, 160
 - Factor
 - Observed, 160
 - Unobserved (latent), 160
 - Linear combination, 160
- Complexity, 99
- Compression algorithms
 - Lossless
 - Arithmetic coding, 295
 - Huffman coding, 324
 - Lossy
 - JPEG (image), 165
- Conceptual models, 21
 - Entity-relationship model (ER), 25
 - Extended entity-relationship model (EER), 25
 - Resource description framework (RDF), 21, 30
 - Object, 111
 - Predicate, 111
 - Properties, 30
 - Reification, 31
 - Resources, 30
 - Statements, 30
 - Subject, 111
 - Unified modeling language (UML), 21, 25
- Content analysis
 - Web page, 225
- CUR decomposition, 177
 - Sampling
 - Biased subspace sampling, 178
 - Uniform sampling, 177
- Database system, 20
 - Components
 - Classification, *see* Classification, 14
 - Clustering, *see* Clustering, 14
 - Data registration, 13
 - Feature extraction, 13, 14
 - Index structures, 13
 - Index structures, *see* Index structures
 - Query interface, 13, 15
 - Query processing, 13, 16
 - Query processing, *see* Query, 13
 - Query refinement, 18
 - Ranking, 17
 - Ranking, *see* Query, 13
 - Recommendation, 15
 - Recommendation, *see* Recommendation systems, 13
 - Relevance feedback, 17
 - Relevance feedback, *see* Relevance feedback, 13
 - Result visualization, 13
- Data model, 14
 - Conceptual, *see* Conceptual models, 21
 - Content based, 20
 - Logical, *see* Logical models, 21
 - Metadata based, 20
 - Physical, *see* Physical models, 21
 - Schema, 20
- Digital library, 112
- Dimensionality curse, 143–145, 180, 263, 264, 267, 268, 341, 374, 380
- Dimensionality reduction, 145, 148, 165, 167
 - Intrinsic dimensionality
 - Correlation fractal dimension, 156
 - Correlation, 153
 - Database compactness, 161
 - Feature selection, *see* Feature selection, 148
 - Intrinsic dimensionality, 153, 159
 - Transformations, *see* Vector space transformations, 156
- Dirichlet process
 - Base distribution, 139
 - Chinese restaurant process, 140
 - Concentration parameter, 139
 - Probability distribution, 139
- Discrete cosine transform (DCT), 90, 163–165
 - 2D-DCT, 164
- Discrete Fourier transform (DFT), 164, 218
- Discrete wavelet transform (DWT), 43, 163, 165
- Discrete wavelet transform (DWT), *see* Wavelet functions, 43

476 Index

- Distance/similarity measures, 9, 11, 102
 - Graph model
 - Edit distance, *see* Graph edit distance, 209
 - Histogram similarity, 209, 231
 - Vector encoding similarity, 210
 - Information theoretic
 - Information distance, 201
 - Normalized compression distance, 201
 - Normalized information distance, 201
 - Probabilistic model
 - Jensen-Shannon divergence, 108
 - KL distance, *see* Kullback-Leibler (KL) divergence, 293
 - Pearson's Chi-square test, *see* Pearson's Chi-square test, 108
 - Properties
 - Metric, *see* Metric distance measures, 33
 - Symmetric, 33, 108
 - Triangular equality, 33, 170
 - Semantic
 - CP/CV, 221
 - Sets
 - Dice similarity, 105
 - Jaccard similarity, 105, 201
 - Signals
 - Signal-to-noise ratio (SNR), 108
 - Spatial model, *see* Spatial distance/similarity, 74
 - String model
 - Compression distance, 201, 218, 219
 - Edit distance, *see* String edit distance, 214
 - Fingerprint distance, *see* String fingerprinting, 198
 - Ziv-Merhav cross-parsing, 202, 218, 219
 - Temporal model, *see* Temporal distance/similarity, 68
 - Tree model
 - Alignment distance, 212, 216
 - Bottom-up edit distance, 217
 - Cousin set similarity, 218
 - Edit distance, *see* Tree edit distance, 212
 - Inclusion distance, 212, 216
 - Isolated-subtree edit distance, 216
 - Path set similarity, 218
 - String encoding similarity, 218
 - Top-down edit distance, 216
 - Vector encoding similarity, 220
 - Vector model
 - Cosine similarity, 104, 222
 - Dot product (scalar product) similarity, 104, 117
 - Intersection similarity, 105, 222
 - Mahalanobis distance, 105
 - Minkowski distance, 69, 103
 - Minkowski distance, *see* p -norms, 103
 - Quadratic distance, 107
- Edges
 - Detection, 46, 50
 - Canny operator, 43
 - Sobel operator, 43
 - Gradient, 44, 45, 52
 - Histogram, 44
 - Histogram, *see* Histogram, 43
 - Directionality, 43
 - Frequency, 43
 - Linking, 46
- Effectiveness, 380
 - H_β measure, 383
 - Arithmetic mean of precision and recall, 381
 - Coverage, 390
 - F-measure, 382
 - False hit, 380
 - Graded ground truth, 386
 - Harmonic mean of precision and recall, 381
 - Kendall-Tau rank coefficient, 387
 - Maximum and minimum F-measures, 388
 - Miss, 380
 - Noninterpolated average-precision (NIAP), 386
 - Normalized discounted cumulative gain (NDCG), 388
 - Normalized modified retrieval rank (NMRR), 389
 - Novelty, 390
 - Pearson's correlation, 387
 - Precision, 381
 - Precision-recall curve, 384
 - Area under, 385
 - R-precision, 385, 389
 - Rank first relevant, 386, 388
 - Ranked ground truth, 386
 - Ranked results, 383
 - Recall, 381
 - Recall at Θ precision, 386
 - Receiver operator characteristic (RoV) curve, 384
 - Relevance feedback, 390
 - Spearman's rank correlation coefficient, 387
 - Statistical significance, 390
 - One-way ANOVA, 393
 - P-value, 392, 393
 - T-test, 391
 - Two-step sampling, 396
 - Two-way ANOVA, 394
 - U-test, 393
 - Wilcoxon signed-rank test, 392
- Eigen analysis, 230, 277
 - Adjacency matrix, 277
 - First eigenvector (portion of time), 278
 - Second eigenvector (proximity), 278
 - Eigen decomposition, 156, 172
 - Eigenvalue, 232
 - Singular value, 172, 173
 - Eigenvector, 157, 179, 226
 - First, 141, 226
 - Left, 157
 - Right, 157
 - Second, 141
 - Eigenvectors, 224, 277
 - Singular vector, 172, 173, 278
 - HITS, 224
 - Laplacian matrix, 277
 - Second eigenvector (connectedness), 278
 - Latent semantic indexing (LSI), *see* Latent semantic indexing (LSI), 222
 - Principal component analysis (PCA), *see* Principal component analysis (PCA), 156

- Singular value decomposition (SVD), *see* Singular value decomposition (SVD), 292
- Topic distillation, 224
- Transition matrix, 226
- Embeddings, 271
 - Distance-based, 167, 168
 - FastMap, *see* FastMap, 170
 - Multidimensional scaling (MDS), *see* Multidimensional scaling (MDS), 168
- Graphs
 - Spreading activation, 227, 228
- Strings
 - String kernels, 200
- Taxonomies
 - Concept propagation/concept vector (CP/CV), 221, 227
- Tensor-based
 - Tensor decomposition, *see* Tensor decomposition, 178
- Trees
 - Propagation vectors for trees (PVT), 220
- Vector-based, 172
- Vector-based
 - CUR decomposition, *see* CUR decomposition, 177
 - Dimensionality reduction, *see* Dimensionality reduction, 172
 - Probabilistic latent semantic analysis (PLSA), *see* Probabilistic latent semantic analysis (PLSA), 176
 - Singular value decomposition (SVD), *see* Singular value decomposition (SVD), 172
 - Vector space transformations, *see* Vector space transformations, 172
- Entropy, 150, 273, 296
- Entropy
 - Event, 150
 - Frequency, 150
 - Relative, 108, 202, 293
 - Shannon
 - Information content, 150
 - Uncertainty, 150
- Expectation maximization (EM), 176, 321
- Experiential computing, 109
- Extensible Markup Language (XML), 113
 - Attribute, 28
 - Document type definitions (DTDs), 28
 - Element, 28
- False hit, 148, 163, 184, 190, 238, 243, 296, 380
- False positive, *see* False hit, 148
- FastMap, 170, 271, 286
 - Pivot, 170
 - Projections, 171
- Feature, 32, 143
 - Association, *see* Feature association, 153
 - Distribution, 151
 - Power law, 144, 374
 - Uniform, 144
 - Zipf's law, 146
 - High level, 34
 - Importance/significance, *see* Feature significance, 143
 - Indexing, *see* Index structures, 143
 - Joint distribution, 153
 - Low level, 34, 41, 45
 - Representation
 - Fuzzy, *see* Fuzzy models, 99
 - Graphs, *see* Graph model, 99
 - Probabilistic, *see* Probabilistic models, 99
 - Strings, *see* Sequence/string model, 99
 - Trees, *see* Tree model, 99
 - Vectors, *see* Vector model, 99
 - Selection, 380
 - Selection, *see* Feature selection, 145
 - Sparse, 153
 - Transformation
 - Distance bounding, 148
 - Lossy, 148, 162
 - Overestimating, 148, 162
 - Underestimating, 148, 162
 - Vector space, *see* Vector space transformation, 156
 - Type
 - Audio, 67
 - Audio, *see* Audio models, 90
 - Color, *see* Color models, 99
 - Hierarchical, *see* Tree model, 99
 - Local, *see* Local feature models, 52
 - Semantic, 15
 - Shape, *see* Shape models, 99
 - Space, *see* Spatial models, 99
 - Texture, *see* Texture models, 99
 - Time, *see* Temporal models, 99
- Feature association
 - Co-absence, 153
 - Co-occurrence, 153, 316
 - Correlation, 102, 106, 158, 160, 161
 - Covariance, 106, 156, 160
 - Cross-presence, 153
 - Dependence, 293, 315
 - Independence, 143
 - Linear, 102
 - Statistical, 102, 107, 153
- Joint distribution, 315
- Measures
 - ϕ -coefficient, 153
 - Added value, *AV*, 153
 - Certainty factor, *F*, 153
 - Collective strength, *S*, 153
 - Confidence, *c*, 153
 - Conviction, *V*, 153
 - Cosine, 153
 - Gini index, *G*, 153
 - Goodman-Kruskal's λ , 153
 - H-measure, negative correlation, *H*, 153
 - J-measure, *J*, 153
 - Jaccard, ζ , 153
 - Kappa, κ , 153
 - Kloggen, *K*, 153
 - Laplace, *L*, 153
 - Monotonic, 153
 - Mutual information, *see* Mutual information, 153
 - Null invariance, 153
 - Odd's ratio, α , lift, 153
 - Piatetshy-Shapiro's, *PS*, 153
 - Support, *s*, 153

478 Index

- Feature association (*cont.*)
 - Symmetric, 153
 - Yule's Q , 153
- Semantic, 102
- Feature selection, 102, 145, 148, 264, 301
 - Dimensionality reduction, *see* Dimensionality reduction, 145
- Feature significance, *see* Feature significance, 145
- LASSO, 303
- Relevance feedback, 400
- Support vector machines, 305
- Feature significance, 143, 145
 - Application requirements, 32
 - Database compactness-based, 161
 - Discrimination, 32, 145, 148, 151, 190
 - Data distribution, 151
 - Entropy, 151
 - Fisher's discriminant ratio, *see* Fisher's discriminant ratio, 151
 - Variance, 151
 - Information theoretic, 150
 - Entropy, 150
 - Frequency, 150
 - Information gain, *see* Information gain, 151
 - Object description, 146
 - Perception, 32, 145
 - Query description, 32, 146
 - User preference, 32
 - Variance, 106
- Field-based spatial relationships
 - Directional
 - Nine directional lower-triangular (9DLT) matrix, 77
 - Topological
 - Nine intersection matrix, 78
 - UID-matrix, 77
- Fisher's discriminant ratio, 151, 299, 301, 391
- Fractals, 240
 - Contraction mapping, 240
 - Hausdorff dimension, 240
 - Hilbert, 240, 241, 259
 - Iterated function system, 43, 240
 - Peano-Hilbert, 241
 - Scale-free, 231
 - Self-similarity, 43, 240
 - Z-order, 242
- Fuzzy criterion
 - Relative importance, 120, 121
 - Partial derivative, 123
- Weighting
 - Fagin's generic weighting function, 121
 - Weighted arithmetic average, 121
 - Weighted product, 123
- Fuzzy models, 10, 34, 99, 115
 - Crisp set, 115
 - Fuzzy logic, 115
 - Fuzzy criterion, *see* Fuzzy criterion, 120
 - Fuzzy score, 116
 - Operators, *see* Fuzzy operators, 116
 - Type-1 predicate, 116
 - Type-2 predicate, 116
 - Fuzzy Petri nets, 65
 - Fuzzy set, 115
 - Membership function, 115
- Fuzzy operators
 - Average score, 118
 - Average semantics, 116, 339
 - Arithmetic, 117
 - Geometric, 117
 - Merge function, 116
 - Monotonic, 339
 - Min semantics, 116, 117, 339
 - Product semantics, 116, 117, 339
 - Score distribution, 119
 - Strong α -cut, 196
 - Triangular-norms (t-norms) and conorms (t-conorms), 116
 - Associativity, 116
 - Boundary condition, 116
 - Commutativity, 116
 - Monotonicity, 116
- Gaussian (normal) distribution, 53, 116, 125, 160
 - Anderson-Darling test, 286
 - Central limit theorem, 116, 125
 - Kolmogorov-Smirnov test, 160
 - Kurtosis test, 160
 - Negative entropy test, 160
- Generative models
 - Collective model, 138
 - Dirichlet models
 - Dirichlet process, *see* Dirichlet process, 139
 - Infinite mixture model, 140
 - Rich-gets-richer phenomenon, 140
 - Stick-breaking construction, 140
- Language models, *see* Language models, 137
- Model estimation
 - Markov-chain based, 138
 - Object model, 138
 - Query model, 138
 - Random-walk based, 138
- Object model, 137
- Object relevance, 138
- Query model, 137
- Result imprecision, 138
- Graph edit distance, 114, 208
 - Edit operations, 114
 - Edge deletion, 209
 - Edge insertion, 209
 - Edge update, 210
 - Vertex deletion, 209
 - Vertex insertion, 209
 - Vertex update, 209, 210
- Graph matching, 208, 210
 - Alignment, 210, 217
 - Common subgraphs, 114
 - Edit distance, 209
 - Edit distance, *see* Graph edit distance, 208
 - Fingerprint, 209
 - GraphGrep, 208, 217
 - Histogram, 208, 209, 217
 - Isomorphism, 114
 - Mapping (correspondence), 210
 - Probe, 209, 217
- Graph model, 33, 99, 111, 208, 227
 - Acyclic, 115
 - Cycle detection, 113

- Adjacency matrix, 222, 275, 277, 278
- Applications
 - Citations, 111
 - Media composition, 111
 - Media references, 111
 - Semantic relationships, 111
 - Social relationships, 112
 - Spatiotemporal distribution, 111
 - Web links, 111, 208
- Bipartite, 229
- Clustering, *see* Clustering, 230
- Conceptual
 - Resource description framework (RDF), *see* Resource description framework (RDF), 30
- Connected, 115
- Edge, 111
 - Directed, 111
 - Labeled, 111
 - Nonlabeled, 111
 - Undirected, 111
- Laplacian matrix, 275, 277, 278
- Link analysis, *see* Link analysis, 222
- Node, 111
 - Labeled, 111, 208
 - Nonlabeled, 111
 - Vertex, 111
- Path, 112, 208
 - ID-path, 208
 - Label-path, 209
 - Shortest, 113, 210, 232, 233
- Probabilistic, *see* Bayesian networks, 136, *see* Markovian models, 140, *see* Hidden Markov Models (HMM), 142
- Problems
 - All-pairs shortest path, 210
 - Connected components, 113
 - Matching, *see* Graph matching, 114
 - Maximal cliques, 114
 - Minimum spanning tree, 113
 - Others, 114
 - Reachability, 113
 - Shortest-path, 113
 - Steiner tree, 113
- Proximity search, *see* Proximity search, 232
- Scale-free, 231
- Spatial, *see* Spatial models, 79
- Trees, *see* Tree model, 212
- Undirected, 115, 208
- Vertex degree (valence), 209, 210, 216, 231–233
- Graph models
 - Temporal, *see* Temporal models, 63
- Heterogeneity, 1, 3, 4, 92
- Hidden Markov models (HMM), 316
 - Backward algorithm, 320
 - Baum-Welch method, 320
 - Expectation maximization (EM), 321
 - Classification, 316
 - Forward algorithm, 318
 - Forward-backward algorithm, 318
 - Hidden states, 317
 - Learning parameters, 320
 - Observation sequence, 318
 - Pattern recognition, 142, 316
 - Predicting the sequence of hidden states, 319
 - Probability of an observation sequence, 318
 - Training, 142, 316
 - Viterbi algorithm, 319
- Histogram, 33, 90, 100, 149, 208, 209, 217, 231, 374
 - Bin, 40, 107
 - Vector representation, *see* Vector model, 40
- Hough transform, 50, 83
 - Accumulator matrix, 51
 - Circle, 52
 - Histogram, 52
 - Histogram, *see* Histogram, 52
 - Line, 50
 - Voting, 50
- Human eye
 - Cones (R,G,B), 34
 - Rods, 34
 - Sensitivity, 32, 34, 36, 37, 40
- Hypermedia document, 109
- Image representation, 21
 - Color, *see* Color models, 40
 - Nonspatial, 40
 - Feature locales, 75
 - Feature, *see* Feature, 76
 - Tile, 76
- Hierarchical models, 62
- Hierarchical models, *see* Tree model, 62
- Object, 10, 62
- Salient points, *see* Local feature models, 52
- Segments, *see* Shape models, 76
- Shape, *see* Shape models, 45
- Spatial models, 10, 63
- Spatial models, *see* Spatial models, 74
- Texture, *see* Texture models, 43
- Wavelet, *see* Discrete wavelet transform (DWT), 43
 - Multiscale, 43
- Imprecision, 8, 10, 65, 93, 99, 116, 123, 126, 138, 311, 379, 398, 399, 409
 - Statistical, 8, 123, 126
- Index structures, 9
 - Graph/tree indexing, *see* Graph model, 208
 - Hierarchical, 235
 - Hierarchical
 - B-trees, B+-trees, 182, 236, 244, 255, 258, 267, 349, 365, 366
 - Balanced, 182, 203, 255, 258
 - Fanout, 182, 252, 253, 258, 262, 264, 267, 375
 - Multidimensional, *see* Multidimensional index structures, 235
 - Sequence/string indexing, *see* Sequence/string model, 181
 - Space-filling curves, *see* Space-filling curves, 238
 - Space subdivision, 235, 244
- Information gain, 151, 298, 301
 - Entropy
 - C4.5, 298
 - C5.0, 298
 - ID3, 298
 - Gini impurity, 298
- Information retrieval (IR), 99, 117, 136, 223

480 Index

- Instant-based temporal relationships
 - =, 58
 - After, 58
 - Before, 58
- Interval-based temporal relationships
 - Binary
 - =, 60
 - After, 60
 - Before, 60
 - Contains, 60
 - During, 60
 - Finished_by, 60
 - Finishes, 60
 - Meets, 60
 - Met_by, 60
 - Overlapped_by, 60
 - Overlaps, 60
 - Started_by, 60
 - Starts, 60
 - N-ary, 61
- Inverted files, 181, 223
 - Multi-keyword, 183
 - Similarity accumulator, 183
- Optimizations
 - d*-gaps, 183
 - Compressed inverted list, 183
 - Elias code, 183
 - Golomb code, 183
 - Sorted inverted list, 183
 - Postings file, 181, 190
 - Search structure, 182
 - Signature file hybrid, 190
 - Word directory, 181
- Kolmogorov complexity, 201, 324
- Kronecker delta, 101
- Kullback-Leibler (KL) divergence, 108, 222, 293
- Language models, 136
 - Generative models, *see* Generative models, 137
 - Smoothing
 - Background (collection) model, 137
 - Dirichlet, 137
 - HMM-based estimation, 137
 - Linear interpolation, 137
 - Semantic, 137
 - Translation
 - Markov-chain based, 137
- Laplace law of succession, 314
- Latent semantic analysis (LSA), *see* Latent semantic indexing (LSI), 173
 - Probabilistic, *see* Probabilistic latent semantic analysis (PLSA), 176
- Latent semantic indexing (LSI), 173, 222, 277, 292
 - Eigen documents, 173
 - Latent semantic analysis (LSA), 174
 - Latent semantics, 173
 - Latent semantics
 - Concepts, 173
- Matrix
 - Document-document, 173
 - Document-term, 173
 - Term-term, 173
- Singular value decomposition (SVD), *see* Singular value decomposition (SVD), 173
- Link analysis, 222
 - Authorities, 223
 - Bi-partite core law, 232
 - Clustering coefficient, 232
 - Collaborative filtering, *see* Collaborative filtering, 229
 - Diameter, 232
 - Distribution of shortest path lengths, 232
 - HITS, 142, 223, 226
 - Hubs, 223, 231
 - Page-to-page associations, 226
 - Seed pages, 226
 - PageRank, 138, 142, 225
 - Random surfer, 225
 - Power law, 231
 - Social networking, *see* Social networks, 230
 - Spreading activation, *see* Spreading activation, 226
 - Topic distillation, 224
 - Topic drift, 224
 - Triangle law, 231
- Local feature models, 52
 - Invariance
 - Deformations, 53
 - Lighting, 53
 - Projection, 53
 - Rotation, 53
 - Scaling, 53
 - Translation, 53
 - Keypoint, 54
 - Descriptor, 54
 - Local descriptor, 53
 - Others, 54
 - Salient point, 53
 - Scale-invariant feature transform (SIFT), 53
 - Keypoint filtering, 54
 - Orientation assignment, 54
 - Scale-space extrema detection, 53
 - Set-based, 52
- Locality sensitive hashing (LSH), 200, 201, 335
- Logical models, 21
 - Expressive power, 25
 - Object-oriented model
 - Abstract data types, 26
 - Class, 26
 - Complex data types, 26
 - Entity, 25
 - Inheritance, 26
 - Methods (behaviors), 25
 - Object identifier, 25
 - ODMG, 25
 - Relationships, 25
 - Object-relational model, 27
 - SQL3, 28
 - Relational model, *see* Relational model, 21
 - Semi-structured models
 - Extensible Markup Language (XML), *see* Extensible Markup Language (XML), 28
 - Object exchange model (OEM), 28
 - Schemaless, 28
 - Self-describing, 28

- Markovian models, 277, 316
- Markov chain, 123, 136–138, 140, 277
 - Proximity, 141
 - Stationary distribution, 141
 - Timed, 66
- Probabilistic process, 140
- Random walk, 139, 225, 226, 277
- Transition graph, 140
- Transition matrix, 139, 140, 225, 226
 - Aperiodic, 141
 - Irreducible, 141
- Matrix, 178
 - Convolution, 53
 - Decomposition
 - CUR Decomposition, *see* CUR
 - Decomposition, 177
 - Eigen decomposition, 157
 - Probabilistic latent semantic analysis (PLSA), *see* Probabilistic latent semantic analysis (PLSA), 176
 - Singular value decomposition (SVD), *see* Singular value decomposition (SVD), 172
 - Dense, 177
 - Diagonal, 107, 172, 177
 - Identity, 106
 - Orthonormal, 172
 - Real-valued, 172
 - Sparse, 177
 - Square, 156, 172
 - Symmetric, 157, 172
- Maximum likelihood estimation (MLE), 314
 - Small-sample correction, 314
 - Smoothing, 314
- Metric distance measures, 33, 103, 108, 239
 - Non-negative, 33
 - Symmetric, 33
 - Triangular inequality, 33
- Minimum description length (MDL), 204, 295, 323, 324
 - Compression, 324
 - Kolmogorov complexity, 324
- Miss, 148, 163, 243, 296, 380
- Model selection, 322
 - Akaike's information criterion (AIC), 323
 - Bayesian information criterion (BIC), 285, 323
 - Degree of complexity, 322
 - Degree of fit, 322
 - Minimum description length (MDL), *see* Minimum description length (MDL), 323
- Multidimensional scaling (MDS), 168, 170, 210, 271, 275, 292
 - Data-driven pivots, 169
 - Scree test, 168
 - Space-driven pivots, 169
 - Stress, 168
 - Stress minimization, 168
- Multidimensional index structures, 102, 235, 244
 - Adaptive
 - Data dependent, 253
 - Fair-split, 254
 - Space dependent, 253
 - VAMSplit, 254, 262
 - BSP-tree, 254
 - Deletion, 248–251, 253, 255, 256, 258, 260, 263, 267
 - Restructuring, 249
 - Grid files, 245
 - Insertion, 246, 248, 251–253, 255, 256, 258, 260–263, 266, 267
 - KD-trees, 252, 329
 - Adaptive, 253
 - Hybrid-tree, 256
 - k-d-B-tree, 255
 - LSD-tree, 255
 - Matrix (MX), 253
 - Point, 253
 - Point-region (PR), 253
 - Live space, 257
 - Metric space
 - Burkhard-Keller tree, 330
 - Generalized-hyperplane tree (GH-Tree), 330, 332
 - Geometric near-neighbor access tree (GNAT), 330, 332
 - M-tree, 261, 262, 264, 330
 - MVP-tree, 330
 - Post-office tree, 330
 - Vantage-point tree (VP-tree), 330
 - Minimum bounding region (MBR), 258–264, 266, 267, 329, 330, 334, 369
 - Overlap, 257, 259, 260, 262, 267
 - Pyramid-trees, 267
 - Quadtrees, 246, 329
 - Matrix (MX), 250
 - Point, 246
 - Point-region (PR), 251
 - R-trees, 257, 264, 329, 349, 368, 375
 - aR-tree, 372
 - Hilbert, 259
 - Hilbert packed, 263
 - Linear-split, 263
 - Packed, 263
 - Quadratic-split, 263
 - R*-tree, 260, 263, 269
 - R+-tree, 260
 - Sort-Tile-Recursive (STR) packed, 263
 - SR-tree, 261
 - SS-tree, 261, 264
 - TV-tree, 263
 - VAMSplit, 262
 - X-tree, 262, 269
 - Self-clustering, 271
 - Space sub-division, 244
 - Closed, 244
 - Open, 244
 - Rectilinear, 244
 - Split, 244, 250–253, 256–258, 261, 262, 267
 - TV-trees
 - Extension, 267
 - Vector telescoping, 264
 - VA-files, 269
 - VA+-file, 269
- Multimedia authoring, 56
- Synchronization, 59
- User interaction, 59

482 Index

- Multimedia standards
 - MHEG-5, 56
 - QuickTime, 56
- Multiple approximate sequence matching, 206
 - Fingerprint based, 206
 - NFA based, 206
- Multiple sequence matching, 204
 - Approximate, *see* Multiple approximate sequence matching, 206
 - Filter sequence, 204
 - Hash based
 - Bloom filter, 205
 - Karp-Rabin (KR) algorithm, 205
 - Trie based, *see* Trie, 205
- Mutual information, 153, 293, 315
- Nearest neighbor search, 143, 145, 235, 301, 328
 - AESA, 333
 - Approximate, 335
 - Batch, 336
 - Branch and bound, 328, 335
 - Best bin first, 328
 - Best-first, 328
 - Delaunay graphs, 332
 - Euclidean space, 329
 - Hierarchical partitioning, 328
 - Incremental, 334
 - LAESA, 333
 - Locality sensitive hashing (LSH), *see* Locality sensitive hashing (LSH), 335
 - Metric space, 330
 - Nonhierarchical, 332
 - Orchard's algorithm, 332
 - Vector model, 11, 235, 248, 263, 267
 - Voronoi decomposition, 332
- Neural networks, 290
 - Hebbian learning rule, 290
 - Neurons, 290
 - Synaptic connection, 290
- Object complexity, 2, 5, 6, 12, 14, 15, 26, 28, 32, 33, 45, 50, 58, 72, 112
- Object vector, 106
- Outlier, 107
- Pearson's chi-square test, 108
- Physical models, 21
 - Column-oriented models, 25
 - Storage, *see* Index structures, *see* Storage, 25
- Point-based spatial relationships
 - Directional
 - Spatial orientation graph, 79
- Precision, 381
- Prefix search, *see* Trie, 181
- Principal component analysis (PCA), 161, 172, 222, 292
- Principal component analysis (PCA)
 - Dimensionality reduction
 - Error optimal Euclidean subspace, 159
 - Kaiser-Guttman rule, 159
 - Mean eigenvalue, 159
 - Parallel analysis, 159
 - Scree test, 159
 - Variance explained, 159
 - Variance maximizing subspace, 158
- Gaussian assumption
 - Statistical independence, 160
- Karhunen-Loeve (KL) transform, 156
- Principal components, 284
- Principal components
 - Eigenvectors, 157
- Transformation
 - Correlation eliminating, 158
 - Linear, 158
- Priority queue (heap), 229, 283, 328, 334, 335, 357, 358, 369
- Probabilistic databases, 129
 - "Unknown" value, 128
 - Disjoint-independence, 130
 - Domain-independence, 132
 - Possible worlds, 128, 130
 - Probabilistic relations, 128, 130
 - Incomplete, 128
 - Key-value pair, 128
 - Probabilistic or-set-tables (p-or-set-tables), 128
- Query, 129
- Query processing, 130, 136
- Representation
 - Closed, 129
- Results, 129
 - Consensus-based, 130
 - Lineage, 134
 - Mean-answer, 130
 - Median-answer, 130
 - Possible tuples, 130
- Tuple-Independence, 132
 - Ordered binary decision diagrams (OBDDs), 132
- Probabilistic latent semantic analysis (PLSA), 176
 - Aspect model, *see* Aspect model, 176
 - Latent semantic indexing (LSI), *see* Latent semantic indexing (LSI), 176
- Probabilistic model
 - Independence, 130
- Probabilistic models, 10, 34, 99, 123
 - Log-likelihood function
 - Expectation maximization (EM), *see* Expectation maximization (EM), 176
- Bayesian models, *see* Bayesian models, 123
- Conditional (a posteriori) probability, 126, 135, 136, 140
- Correlation, 127
- Covariance, 127
 - Matrix, 107, 156, 172
- Cumulative distribution function, 124
- Databases, *see* Probabilistic databases, 127
- Density function, 124
- Distributions, 108
 - Binomial, 125
 - Exponential, 125
 - Gamma, 125
 - Gaussian (normal) distribution, *see* Gaussian (normal) distribution, 125
 - Geometric, 125
 - Multinomial, 125
 - Negative binomial, 125

- Uniform, 125
- Vernoulli, 125
- Zipfian, 125, 146, 231
- Divergence, *see* Kullback-Leibler (KL) divergence, 293
- Fit, *see* Pearson's Chi-square test, 108
- Generative models
 - Aspect model, *see* Aspect model, 176
- Imprecision, *see* Imprecision, 123
- Independence, 107, 126, 160
- Marginal (prior) probability, 126, 135
- Markovian, *see* Markovian models, 140
- Mass function, 124
- Mean, 124
- Mixture models, 140
- Probability distribution
 - Continuous, 124
 - Discrete, 124
- Random variable, 124
- Stochastic (random, probabilistic) process, 137, 140, 317
- Stochastic (random, probabilistic) process
 - Dirichlet process, *see* Dirichlet process, 139
 - Markovian, *see* Markovian models, 140
- Stochastic Petri nets, 65
- Uncertainty, *see* Probabilistic databases, 127
- Variance, 126, 151, 156
- Procrustes algorithm, 211
 - Orthogonal, 211
- Proximity search, 232
 - Algorithms
 - BANKS-I, BANKS-II, 233
 - DPBF, 233
 - RIU, 233
 - Group Steiner tree, 233
 - Information unit, 233
 - Minimal answer, 232
- Query
 - Processing
 - Probabilistic databases, *see* Probabilistic databases, 130
- Query
 - Fuzzy query, *see* Fuzzy models, 339
 - Keyword query
 - Conjunctive, 185
 - Disjunctive, 187
 - Single, 185
 - Language
 - Multimedia, 92
 - Relational, *see* Relational model, 23
 - Nearest neighbor query, *see* Nearest neighbor search, 235
 - Partial match query, 117
 - Popularity query, 110
 - Prefix search, *see* Trie, 181
 - Probabilistic query, *see* Probabilistic models, 129
 - Processing, 13, 16, 148, 401
 - Optimization, *see* Query Optimization, 373
 - Ranking algebra, *see* Top-K query processing, 23
 - Relational algebra, *see* Relational algebra, 23
 - Sequential scan, 144, 169, 235, 263, 269, 272, 341, 364
 - Top-K query processing, *see* Top-K query processing, 354
 - Proximity query, *see* Proximity search, 232
 - Range query, *see* Range search, 235
 - Relational query, *see* Relational model, 22
 - Skyline query, *see* Skylines, 360
 - Specification
 - Declarative, 373
 - Query by description, 9
 - Query by example (QBE), 9, 15
 - Visual, 10
 - Vector, 106
- Query optimization
 - Cost estimation, 374
 - Cost model, 16, 374, 377
 - Expensive predicates, 374
 - Fanout, 377
 - Filter predicates, 375
 - Join predicates, 377
 - Predicate migration, 375
 - Restriction predicates, 375
 - Fanout model, 377
 - Quality model, 16, 377
 - Query plan, 25, 374
 - Rank aware, 378
 - Ranking predicates, 378
 - Selectivity estimation, 374
 - Power law, 374
 - Spatial histograms, 374
 - Statistics, 16, 374
 - Top-K, 374
- Range search, 11, 143, 144, 161, 169, 189, 235, 355, 375, 383, 401
 - Space-filling curves, 243
 - Vector model, 148, 152, 190, 235, 243, 248, 251, 257, 258, 263, 267, 328, 334, 335
- Ranked join, 338
 - Anytime, 344
 - Combined Algorithm (CA), 344
 - Convex hull, 347
 - Fagin's Algorithm (FA), 339, 355
 - Layer ordering, 346
 - Merge function, 339
 - Average, 339
 - Minimum, 339
 - Monotonic, 339
 - Product, 339
 - No random access algorithm (NRA), 343, 357
 - Probabilistic, 344
 - Variants, 344
 - Onion technique, 347
 - Partial sorted access (PSA), 345
 - RANK-JOIN, 343, 358
 - Relaxed monotonicity, 348
 - Cosine, 352
 - Horizon based ranked join (HR-Join), 351
 - Skip-and-prune (SnP), 353
 - Sum-Max Monotonicity, 349
 - Robust indexing, 347
 - Threshold Algorithm (TA), 341, 357
 - Variants, 342

484 Index

- Ranked join (*cont.*)
 - Top-K query processing in the DBMS, *see* Top-K query processing, 354
- Ranked retrieval, 183, 327
- Recall, 381
- Recommendation systems, 9, 15, 110, 131, 229, 230, 400, 413
- Recommendation systems, *see* Collaborative filtering, 13
- Regular expressions, 202
 - Alphabet, 202
 - Concatenation, 202
 - Finite automaton, 203
 - Finite automaton
 - Deterministic (DFA), 193, 203
 - Minimal bounding, 203
 - Nondeterministic (NFA), 193, 203
 - Index structures
 - RE-tree, 203
 - Kleene star, 202
 - Regular languages, 203
 - Size of, 204
 - Union, 202
- Relational algebra
 - Aggregation operator (Γ), 23, 24
 - Cartesian product operator (\times), 23, 24
 - Join operator (\bowtie), 23, 24
 - Projection operator (π), 23, 24
 - Rename operator (ρ), 23, 24
 - Selection operator (σ), 23, 24
 - Set operators (\cup, \cap, \setminus), 23, 24
- Relational model
 - Closed-world assumption, 21
 - Query languages
 - Declarative, 22
 - Domain relational calculus, 22
 - SQL, 10, 23
 - Tuple relational calculus, 23
 - Query processing
 - Query optimization, *see* Query optimization, 24
 - Relational algebra, *see* Relational algebra, 23
- Relation
 - Extension, 21
 - Tuple, 21
- Schema, 22
 - Attribute, 22
 - Candidate key, 22
 - Foreign key, 22
 - Integrity constraints, 22
 - Referential integrity, 22
- Relationship
 - Nonsymmetric, 111
 - Symmetric, 111
- Relevance feedback, 8, 12, 17, 32, 102, 120, 148, 288, 400
 - Bias in feedback, 401
 - Decay, 411
 - Divergence minimization, 410
 - Feedback drift, 401
 - Generative models, 409
 - Ide dec-hi algorithm, 404
 - Language modeling, 408
 - Negative feedback, 399, 410
 - Ostensive relevance, 411
 - Positive feedback, 399
 - Probabilistic, 123, 136, 404
 - Pseudo-feedback, 411
 - Query rewriting, 404
 - Rocchio's algorithm, 404
 - Vector space, 404
- Relevance filtering, 10
- Retrieval
 - Associative, *see* Spreading activation, 226
 - Content-based, 1, 226
- Scree test, 159, 168
- Semantic gap, 1, 6, 8, 380, 398, 399
- Sensors, 67, 116, 178
- Sequence/string model, 33, 67, 99, 181
 - Alphabet, 192
 - Distance measure
 - Compression distance, 201, 218, 219
 - Edit distance, *see* String edit distance, 218
 - Ziv-Merhav cross-parsing, 202, 218, 219
 - Index structures
 - Inverted files, *see* Inverted files, 181
 - RE-tree, 203
 - Signature files, *see* Signature files, 181
 - Suffix indexes, *see* Suffix indexes, 181
 - Trie, *see* Trie, 181
 - Pattern matching, *see* Regular expressions, 202
 - Prefix matching, *see* Trie, 192
 - Regular expressions, *see* Regular expressions, 202
- Sequence matching, 191
 - Approximate, *see* Approximate sequence matching, 195
 - Boyer-Moore (BM) algorithm, 191, 205
 - Knuth-Morris-Pratt (KMP) algorithm, 191, 205
 - Multiple, *see* Multiple sequence matching, 204
 - Suffix indexes, *see* Suffix indexes, 192
- Wildcards
 - "*" wildcard, 110, 202
 - "/" wildcard, 110, 202
- Shape models, 45, 99
 - Area, 49
 - Boundary, 47
 - B-splines, 48
 - Chain code, 48
 - Piece-wise linear, 48
 - Time series, 48
 - Circularity, 49
 - Connected components, 49
 - Diameter, 49
 - Eccentricity, 50
 - Height, 49
 - Histogram, 49, 52
 - Histogram, *see* Histogram, 49
 - Holes, 49
 - Hough transform, *see* Hough transform, 50
 - Major axis, 50
 - Moment, 50
 - Central, 50
 - Perimeter, 49
 - Segmentation, 76

- Clustering, 46
- Edge linking, 46
- Partitioning, 46
- Region growing, 46
- Watershed transformation, 47
- Spread, 50
- Width, 49
- Signal, Power, 91
 - 1D, 164
 - 2D, 164
 - Amplitude, 91
 - Continuous, 164
 - Cosine decomposition
 - Cosine waves, 164
 - Discrete, *see* Discrete cosine transform (DCT), 164
 - Discrete, 164
 - Domain
 - Frequency, 164
 - Time, 164
 - Fourier decomposition
 - Discrete, *see* Discrete Fourier Transform (DFT), 164
 - Sinusoidal waves, 91
 - Sinusoids, 164
 - Frequency, 164
 - Noise, 165
 - Noise
 - Signal-to-noise ratio (SNR), 108
 - Nonrandom
 - Temporally correlated, 91, 165
 - Random, 165
 - Wavelet decomposition
 - Discrete, *see* Discrete Wavelet Transform (DWT), 165
 - Wavelet functions, *see* Wavelet functions, 165
- Signature files, 181, 184, 205
 - Bitslice, 187, 190
 - Bitwise-or, 184
 - Blocked, 187, 190
 - Document signature, 184
 - False positives, 184, 188
 - Inverted file hybrid, 190
- Query
 - Conjunctive, 185
 - Disjunctive, 187
 - Document query, 189
 - Single keyword, 185
- Word signature, 184
 - Hash function, 184
- Similarity measures, *see* Distance/Similarity measures, 293
- Singular value decomposition (SVD), 172, 173, 292
 - High order, 179
 - Incremental, 174
 - Folding, 174
 - SVD-Update, 174
- Latent semantic indexing (LSI), *see* Latent semantic indexing (LSI), 173
- Skylines, 360
 - B-tree based, 365, 366
 - Bitmap skylines, 367
 - Block-nested-loops skylines algorithm, 361
 - Branch-and-bound based, 368
 - Divide-and-conquer based, 365
 - Linear elimination sort skylines algorithm, 365
 - Maximal vector, 361
 - Nearest neighbor based, 368
 - Nearest neighbor search, *see* Nearest neighbor search, 368
 - Nest-loop based, 361
 - Pareto curve, 361
 - Pareto frontier, 361
 - Partially ordered data, 369
 - l*-cuts, 371
 - Hasse-diagram, 369
 - Interval mapping, 370
 - Lattice, 369
 - Poset, 369
 - Weak Pareto dominance, 371
 - Presorting based, 364
 - Preference function, 361
 - R-tree based, 368
 - Sort-filter-skyline algorithm, 364
 - Top-K dominating queries, 361, 372
 - Window, 362
- Social networks, 67, 99, 141, 230, 231
- Social networks
 - Betweenness, 230
 - Centrality, 230
 - Clustering, *see* Clustering, 230
 - Cohesion, 230
 - Friends, 230
 - Key individuals, 230
 - Authorities, 230
 - Hubs, 230
 - Preferential attachment, 230
 - Prestige, 230
 - Prominence, 230
 - User similarity, 230, 231
- Space-filling curves, 238, 267
 - Cantor-diagonal-order, 239
 - Column-order, 238
 - Fractals, *see* Fractals, 240
 - Hilbert, 240, 241, 259
 - Peano-Hilbert, 241
 - Row-order, 238
 - Row-prime-order, 239
 - Z-order, 242
 - Bit-shuffling, 243
- Space-filling curves, 238
- Spatial distance/similarity, 82
 - Constraint-based, 82
 - Fuzzy evaluation, *see* Fuzzy models, 82
 - Graph-based, *see* Graph matching, 83
 - 2D-PIR, 88
 - SIM_{DTC}, 90
 - SIM_L, 90
 - Spatial orientation graph, 83
 - Probabilistic evaluation, *see* Probabilistic models, 82
- String-based, *see* String edit distance, 84
 - 2D Θ R-string, 85
 - 2D B-string, 87
 - 2D B ϵ -string, 87
 - 2D C+-string, 86
 - 2D C-string, 86

486 Index

- Spatial distance/similarity (*cont.*)
 - 2D E-string, 86
 - 2D G-string, 86, 90
 - 2D Z-string, 87
 - 2D-string, 84
 - Topological similarity, 83
- Spatial models
 - Distance/similarity measures, *see* Spatial distance/similarity, 74
 - Distribution, 99
 - Fields, 75
 - Field functions, 75
 - Field operations, 75
 - Relationships, *see* Field-based Spatial relationships, 75
 - Spatial framework, 75
 - Line, 75
 - Points, 79
 - Relationships, *see* Point-based relationships, *see* Point-based spatial relationships, 79
 - Reference direction, 81, 90
 - Region, 75
 - Regions-to-points
 - Centroid scheme, 80
 - Line sweep method, 80
 - Minimum bounding rectangle scheme, 80
 - Plane sweep, 80
 - Retrieval
 - Exact, 81
 - Similarity-based, *see* Spatial distance/similarity, 82
- Spatial relationships
 - Directional, 75
 - Quantitative, 75
 - Set-based, 75
 - Topological, 75
- Spreading activation, 226
 - Activation control matrix, 227–229
 - Activation level, 227, 229
 - Activation vector, 227–229
 - Branch-and-bound model, 227, 228
 - Constrained leaky capacitor model, 227, 228
 - Linear transformation, 228
 - Hopfield net model, 227, 228
 - Sigmoid transformations, 228
 - Iterative activation, 227
 - Propagation, 227
 - Semantic memory, 227
- Standards
 - Authoring
 - HyTime, 2
 - MHEG, 2
 - SGML, 2
 - SMIL, 2, 6
 - VRML, 3
 - X3D, 3, 6, 13, 27, 99, 112, 208
 - XML, 2
 - Content description
 - MPEG21, 3
 - MPEG7, 3, 6
- Stimulus
 - Perception
 - Weber-Fechner law, 38
- Stop words, 147
- Storage, 13
 - Disk, 182
 - Page, 235, 245, 255, 256, 262–264, 267, 276
- String edit distance, 67, 110, 167, 214, 218
 - Computation
 - Bit-parallelism, 197, 206
 - Column-oriented NFA simulation, 196
 - Diagonal-oriented NFA simulation, 196
 - Dynamic programming, 196
 - Edit operation
 - Deletion, 110, 197
 - Edit cost, 111
 - Insertion, 110, 197
 - Matching, 197
 - Substitution, 110, 197
 - Swap, 110
 - Transposition, 111
 - Types
 - Bubble sort distance, 110
 - Damerau-Levenshtein distance, 111
 - Episode distance, 110
 - Hamming distance, 110
 - Kendall tau distance, 110
 - Levenstein distance, 111
- String encoding of trees
 - Others, 219
 - Prüfer encoding, 219
- String fingerprinting, 198
 - ρ -grams, 167, 181, 198
 - KR algorithm, 198
 - Local fingerprinting, 199
 - Minsampling, 199, 200
 - Modsampling, 199
 - Rolling hash functions, 198
 - Sampling, 199
 - Winnowing, 199
- Counting filter, 198, 206
- Locality sensitive hashing (LSH), *see* Locality sensitive hashing (LSH), 200
- Maximum error rate, 198
- String kernels, 200
 - ρ -spectrum kernel similarity, 200
 - Positive definite, 200
- String model, *see* Sequence/string model, 181
- Structure
 - Hierarchical, 167, 208
 - Object composition, 33
 - Interaction, 14
 - Spatial, 14, 33, 167, 168, 208
 - Temporal, 14, 33
- Subjectivity, 8, 9, 17, 271, 402
- Suffix indexes
 - Suffix array, 193
 - Suffix automaton, 182, 193
 - BDM algorithm, 194
 - Bit parallelism, 194
 - BNDM algorithm, 194
 - Deterministic, 193
 - Directed acyclic word graph, 193
 - Nondeterministic, 193
 - Suffix tree, 181, 192, 199, 202
- Supervised learning, 297
- Support vector machines, 301, 303
 - Complexity, 307

- Convex hull, 305
- Feature selection, 305
- Hyperplane search, 304
- Kernel parameters, 303
- L1-, 305
- L2-, 305
- Lagrangian formulation, 305
- Linearly separable, 304
- N-ary, 307
- Nonlinear kernels, 306
 - Gaussian radial basis, 306
 - Polynomial, 306
 - Sigmoid, 306
- Nonlinearly separable, 306
 - Kernel methods, 306
 - Soft-margin, 306
- Quadratic optimization, 305
- Support vectors, 305
- Symbolic aggregate approximation (SAX), 68

- Taxonomy, 99, 113, 221
 - Concept, 113, 221
 - Concept similarity
 - CP/CV, 221
- Temporal distance/similarity
 - Constraint relaxation, 72
 - Edit distance, *see* String edit distance, 69
 - Dynamic time warping (DTW), 69
 - Extended timeline, 71
 - Probabilistic, 72
 - Fuzzy evaluation, *see* Fuzzy models, 68
 - Mapped objects, 69
 - Object priorities, 69
 - Timeline, 69
- Temporal models, 55
 - Actions, 55
 - Clock, 65
 - Concurrency, 63
 - Constraint-based, 55
 - Causal models, 59
 - Difference constraints, 58
 - Event calculus, 59
 - Logical constraints, 58
 - Situation Calculus, 59
 - Distance/similarity measures, *see* Temporal distance/similarity, 68
 - Events, 55, 99, 109
 - Graph based
 - Time flow graph (TFG), 65
 - Timed automaton, *see* Timed automaton, 65
 - Timed Petri nets, *see* Timed Petri nets, 63
 - History, 67
 - Hybrid, 62
 - Instant-based, 58
 - Relationships, *see* Instant-based temporal relationships, 58
 - Interval-based, 58
 - Relationships, *see* Interval-based temporal relationships, 60
 - Object composition Petri nets (OCPN), 111
 - Processes, 61
 - Temporal pattern, 65
 - Time dimension, 178
 - Time Series, 67
 - Timeline models, 55
 - Trace, 65
- Tensor, 67, 178
 - Decomposition, *see* Tensor decomposition, 178
 - Mode, 178
 - Order, 178
 - Slice, 178
- Tensor decomposition, 178
 - Diagonal
 - CANDECOMP, 179
 - PARAFAC, 179
 - Incremental
 - Dynamic tensor analysis (DTA), 180
 - Streaming tensor analysis (STA), 180
 - Window-based tensor analysis (WTA), 180
 - Optimal, 179
 - Orthonormal
 - High-order SVD, 179
 - Tucker, 179
- Texture models, 99
- Texture models
 - Directionality, 40, 43
 - Edges
 - see* Edges, 43
 - Edginess, 44
 - Fractals, 42
 - Fractals, *see* Fractals, 42
 - Frequency, 43
 - Granularity, 40, 44
 - Histogram, 43
 - Histogram, *see* Histogram, 43
 - Pattern, 40
 - Periodicity, 40
 - Random fields, 42
 - Generative models, *see* Generative models, 42
 - Stochastic process, *see* Probabilistic models, 42
 - Rate of change, 40
 - Smoothness, 40, 44
 - Wavelets, 43
 - Wavelets, *see* Wavelet functions, 43
- TF-IDF, 147, 284, 327
 - Inverse Document Frequency (IDF), 147
 - Term Frequency (TF), 147
- Timed automaton, 65
 - Construction
 - logO, 65
 - Finite automaton, *see* Regular expressions, 65
- Timed Petri nets
 - Boundedness, 63
 - Dynamic timed Petri nets (DTPN), 65
 - Fuzzy-timing Petri-net for multimedia synchronization (FTNMS), 65
 - Liveliness, 63
 - Object composition Petri nets (OCPN), 64
 - Places, 63
 - Reachability, 63
 - Safety, 63
 - Stochastic Petri nets, 65
 - Tokens, 63
 - Transitions, 63

488 Index

- Top- k query processing, 354
 - J^* join operator, 357
 - Breaking, 356
 - Filter-based, 354
 - HRJN operator, 358
 - Nearest neighbor queries, *see* Nearest neighbor search, 360
 - NRA-RJ join operator, 358
 - Optimization, *see* Query Optimization, 374
 - RANK-JOIN, 358
 - Ranked join, *see* Ranked join, 354
 - Ranking algebra, 358
 - Merge operator, 359
 - Rank operator, 360
 - Ranked relation, 360
 - Ranking predicate, 360
 - Ranking expressions, 354
 - Skyline queries, *see* Skylines, 360
 - SQL extensions, 354
 - Stop and restart, 355
 - Top- k dominating queries, 361, 372
- Top- k retrieval, 327, 337
- Tree edit distance, 213, 214
 - Bottom-up, 217
 - Edit distance, 212
 - Edit operations, 212, 215
 - Node deletion, 215
 - Node insertion, 215
 - Node relabeling, 215
 - Isolated-subtree, 216
 - Top-down, 216
- Tree matching
 - Alignment, 217
 - Alignment distance, 212, 216
 - Edit distance, *see* Tree edit distance, 212
 - Filtering, 217
 - Cousin set similarity, 218
 - Path set similarity, 218
 - Propagation-vectors for trees (PVT), 220
 - String encoding, *see* String encoding of trees, 218
 - Time series encoding, 218
 - Inclusion distance, 212, 216
- Tree model, 33, 99, 112, 212
 - Applications
 - Discussion boards, 113
 - Extensible Markup Language (XML), *see* Extensible Markup Language (XML), 113
 - Hierarchically structured multimedia objects, 112
 - Navigation hierarchy, 113
 - Semi-structured data, 113
 - Taxonomy, *see* Taxonomy, 113
 - Edge
 - Labeled, 213
 - Matching, *see* Tree matching, 212
 - Node
 - Ancestor, 112
 - Child, 112
 - Descendant, 112
 - Labeled, 115, 213
 - Parent, 112
 - Root, 112
 - Sibling, 112
 - Vertex, 112
 - Ordered, 112, 115, 213
 - Rooted, 112, 115
 - Traversal
 - Inorder, 218
 - Postorder, 213, 218
 - Preorder, 218
 - Unordered, 112, 115, 212, 215
- Trees in graphs
 - Group Steiner tree, 233
 - Maximum spanning tree, 279
 - Maximum spanning tree, 316
 - Min-cut tree, 279
 - Minimum spanning tree, 113, 233, 279
 - Spanning tree, 114, 332
 - Steiner tree, 113
- Trie, 181, 182, 192, 243
 - Multiple sequence
 - Aho-Corasick trie, 205
 - Commentz-Walter trie, 205
- Unsupervised learning, 297
- User experience, 109
 - Event sequence, 109
 - Navigation, 109
- Vector model, 33, 99, 221, 227
 - Data distribution, 105, 106
 - Local, 107
 - Power law, 144, 374
 - Uniform, 144
 - Index structures, *see* Multidimensional index structures, 235
 - Nearest neighbor search, 11, 235, 248, 263, 267
 - Nearest neighbor search, *see* Nearest neighbor search, 11
 - Range search, 148, 152, 190, 235, 243, 248, 251, 257, 258, 263, 267, 328, 334, 335
 - Range search, *see* Range search, 148
 - Space, *see* Vector space, 158
- Vector space, 100, 167
 - Addition, 100
 - Alignment
 - Procrustes algorithm, 211
 - Basis vectors (Bases, Dimension vectors), 100, 101, 117, 158, 161, 165–167, 173
 - Weighting, 106
 - Weighting, *see* Feature Significance, 106
 - Collection of elements, 100
 - Dimensionality, 153, 159, 168, 170
 - Curse, *see* Dimensionality curse, 143
 - Reduction, *see* Dimensionality reduction, 145
 - Distance, *see* Distance/Similarity measures, 102
 - Inner product, 101
 - Multiplication, 100
 - Norm
 - p -norms, *see* p -norms, 101
 - L1 (Manhattan), 265, 303, 305, 329, 369
 - L2 (Euclidean), 144, 158, 163, 190, 265, 272, 303, 305, 329
 - Transformation, *see* Vector space transformation, 158
- Vectors
 - Energy, 163

- Length, 101
- Linearly independent, 100, 101, 158, 165, 173
- Nonorthogonal, 178
- Orthogonal, 101, 158, 166, 173
- Orthonormal, 101, 158, 172
- Unit length, 165
- Vector space transformations
 - Angle preserving, 158, 163, 165
 - Components, 164
 - Distance preserving, 158, 163, 165
 - Energy preserving, 163
 - Fixed basis
 - Discrete cosine transform (DCT), *see* Discrete cosine transform (DCT), 163
 - Discrete Fourier transform (DFT), *see* Discrete Fourier transform (DFT), 164
 - Discrete wavelet transform (DWT), 163
 - Invertible transforms, 164
 - Linear transforms, 164
 - Lossy compression, 165, 167
 - Orthonormal, 158, 163
 - Variable basis
 - Common factor analysis (CFA), *see* Common factor analysis (CFA), 160
 - CUR decomposition, *see* CUR decomposition, 177
 - Independent component analysis (ICA), 160
 - Karhunen-Loeve (KL) transform, *see* Principal component analysis (PCA), 156
 - Probabilistic latent semantic analysis (PLSA), *see* Probabilistic latent semantic analysis (PLSA), 176
 - Singular value decomposition (SVD), *see* Singular value decomposition (SVD), 174
 - Tensor decomposition, *see* Tensor decomposition, 178
- Video representation
 - Action, 63
 - Frame, 63
 - Frame, *see* Image representation, 63
 - Sequence, 63
 - Key frame, 63
 - Motion, 63
 - Object, 62
 - Temporal, *see* Temporal models, 63
- Wavelet functions, 165
 - Family of wavelets, 165
 - Haar wavelets, 166
 - Mother wavelet, 43, 165
 - Scaling/dilation parameter, 165
 - Shift/translation parameter, 165
- Web search, 67
- Web search, *see* Link analysis, 222
- Hyperlinks, 222
- IR techniques, 222
- Keyword search, 222
- Page-to-page associations, 222, 226