

Index

- A-Priori Algorithm, 201, 202, 208
- Accessible page, 178
- Active learning, 421
- Ad-hoc query, 126
- Adjacency matrix, 344
- Adomavicius, G., 324
- Advertising, 16, 110, 194, 267
- Adwords, 276
- Affiliation-Graph model, 352
- Afrati, F.N., 66, 382
- Agglomerative clustering, *see* Hierarchical clustering
- Aggregation, 31, 35
- Agrawal, R., 226
- Alon, N., 152
- Alon–Matias–Szegedy Algorithm, 138
- Amplification, 95
- Analytic query, 50
- AND-construction, 95
- Anderson, C., 324
- Andoni, A., 122
- ANF, *see* Approximate neighborhood function
- ANF Algorithm, 376
- Apache, 22, 67
- Approximate neighborhood function, 376
- Arc, 367
- Archive, 124
- Ask, 183
- Association rule, 195, 196
- Associativity, 25
- Attribute, 30
- Auction, 279
- Austern, M.H., 67
- Authority, 182, 183
- Average, 135
- B-tree, 265
- Babcock, B., 153, 265
- Babu, S., 153
- Backstrom, L., 382
- Bag, 38, 71
- Balance Algorithm, 278
- Balazinska, M., 67
- Band, 83
- Bandwidth, 20
- Basket, *see also* Market basket, 191, 193, 194, 222
- Batch gradient descent, 445
- Batch learning, 421
- Bayes net, 4
- BDMO Algorithm, 257
- Beer and diapers, 196
- Bell, R., 324
- Bellkor’s Pragmatic Chaos, 295
- Berkhin, P., 190
- Berrar, D.P., 414
- Betweenness, 333
- BFR Algorithm, 241, 244
- BFS, *see* Breadth-first search
- Bi-clique, 339
- Bid, 277, 279, 285, 286
- BigTable, 66
- Bik, A.J.C., 67
- Binary Classification, 416
- Biomarker, 195
- Bipartite graph, 273, 329, 339, 340
- BIRCH Algorithm, 265
- Birrell, A., 67
- Bitmap, 209
- Block, 11, 19, 170, 171
- Blog, 178
- Bloom filter, 131, 207
- Bloom, B.H., 153
- Blum, A., 458
- Bohannon, P., 67
- Boldi, P., 382
- Bonferroni correction, 5
- Bonferroni’s principle, 4, 5
- Bookmark, 176
- Boral, H., 383
- Borkar, V., 66
- Bottou, L., 458
- Bradley, P.S., 265
- Breadth-first search, 333
- Brick-and-mortar retailer, 194, 292, 294
- Brin, S., 190
- Broad matching, 279

- Broder, A.Z., 18, 122, 190
 Bu, Y., 67
 Bucket, 9, 129, 143, 147, 207, 257
 Budget, 277, 284
 Budiu, M., 67
 Burges, C.J.C., 458
 Burrows, M., 67
 Candidate itemset, 204, 217
 Candidate pair, 83, 208, 210
 Carey, M., 66
 Categorical feature, 416, 455
 Centroid, 230, 233, 238, 241, 245
 Chabbert, M., 324
 Chandra, T., 67
 Chang, F., 67
 Characteristic matrix, 76
 Charikar, M.S., 122
 Chaudhuri, S., 122
 Checkpoint, 43
 Chen, M.-S., 226
 Child, 333
 Cholera, 3
 Chronicle data model, 152
 Chunk, 22, 217, 244
 CineMatch, 321
 Classifier, 303, 415
 Click stream, 125
 Click-through rate, 271, 277
 Clique, 338
 Cloud computing, 15
 CloudStore, 22
 Cluster computing, 19, 20
 Cluster tree, 252, 253
 Clustera, 38, 66
 Clustering, 3, 16, 228, 310, 325, 330, 415
 Clustroid, 232, 239
 Collaboration network, 328
 Collaborative filtering, 4, 16, 70, 267, 292, 306, 328
 Column-orthonormal matrix, 397
 Combiner, 25, 168, 170
 Communication cost, 20, 44, 365
 Community, 16, 325, 336, 338, 361
 Community-affiliation graph, 352
 Commutativity, 25
 Competitive ratio, 16, 272, 275, 279
 Complete graph, 339, 340
 Compressed set, 245
 Compute node, 19, 20
 Computer game, 299
 Computing cloud, *see* Cloud computing
 Concept, 399
 Concept space, 404
 Confidence, 195, 196
 Content-based recommendation, 292, 296
 Convergence, 426
 Cooper, B.F., 67
 Coordinates, 228
 Cortes, C., 458
 Cosine distance, 89, 99, 298, 302, 404
 Counting ones, 142, 257
 Covering an output, 56
 Craig's List, 267
 Craswell, N., 290
 Credit, 334
 Cristianini, N., 458
 Cross-Validation, 420
 Crowdsourcing, 422
 CUR-decomposition, 384, 406
 CURE Algorithm, 249, 252
 Currey, J., 67
 Curse of dimensionality, 230, 254, 452, 455
 Cut, 343, 344
 Cyclic permutation, 81
 Cylinder, 11
 Czajkowski, G., 67
 DAG, *see* Directed acyclic graph
 Darts, 132
 Das Sarma, A., 66
 Dasgupta, A., 382
 Data mining, 1
 Data stream, 15, 220, 256, 270, 434
 Data-stream-management system, 123
 Database, 15
 Datar, M., 122, 153, 265
 Datar–Gionis–Indyk–Motwani Algorithm, 143
 Dead end, 158, 161, 162, 184
 Dean, J., 67
 Decaying window, 148, 222
 Decision tree, 17, 303, 419, 420, 456
 Deerwester, S., 414
 Degree, 341, 362
 Degree matrix, 345
 Dehnert, J.C., 67
 del.icio.us, 299, 329
 Deletion, 90
 Dense matrix, 28, 406
 Density, 238, 240
 Depth-first search, 374
 Determinant, 386
 DeWitt, D.J., 67
 DFS, *see* Distributed file system
 Diagonal matrix, 397
 Diameter, 237, 240, 368
 Diapers and beer, 194
 Difference, 31, 34, 38
 Dimension table, 50
 Dimensionality reduction, 16, 312, 384, 453
 Directed acyclic graph, 333
 Directed graph, 367
 Discard set, 244

- Disk, 11, 199, 230, 252
 Disk block, *see* Block
 Display ad, 267, 269
 Distance measure, 87, 228, 331
 Distinct elements, 133, 137
 Distributed file system, 19, 21, 192, 198
 DMOZ, *see* Open directory
 Document, 69, 72, 194, 229, 286, 297, 299, 418
 Document frequency, *see* Inverse document frequency
 Domain, 181
 Dot product, 89
 Drineas, P., 414
 Dryad, 66
 DryadLINQ, 66
 Dual construction, 330
 Dubitzky, W., 414
 Dumais, S.T., 414
 Dup-elim task, 40
e, 12
 Edit distance, 90, 92
 Eigenpair, 385
 Eigenvalue, 158, 346, 384, 395
 Eigenvector, 158, 346, 384, 389, 394
 Email, 328
 Energy, 402
 Ensemble, 303, 456
 Entity resolution, 104
 Equijoin, 31
 Erlingsson, I., 67
 Ernst, M., 67
 Ethernet, 19, 20
 Euclidean distance, 87, 101, 452
 Euclidean space, 87, 91, 228, 230, 233, 249
 Exponentially decaying window, *see* Decaying window
 Extrapoliation, 450
 Facebook, 16, 176, 326
 Fact table, 50
 Failure, 20, 26, 39, 40, 42
 Faloutsos, C., 383, 414
 False negative, 83, 93, 216
 False positive, 83, 93, 132, 216
 Family of functions, 94
 Fang, M., 226
 Fayyad, U.M., 265
 Feature, 252, 297, 298
 Feature selection, 421
 Feature vector, 416, 455
 Fetterly, D., 67
 Fikes, A., 67
 File, 21, 198, 215
 Filtering, 130
 Fingerprint, 107
 First-price auction, 279
 Fixedpoint, 96, 182
 Flajolet, P., 153
 Flajolet–Martin Algorithm, 134, 376
 Flow graph, 39
 Fortunato, S., 382
 Fotakis, D., 382
 French, J.C., 265
 Frequent bucket, 208, 209
 Frequent itemset, 4, 192, 201, 204, 340, 415
 Frequent pairs, 202
 Frequent-items table, 203
 Freund, Y., 458
 Friends, 326
 Friends relation, 49
 Frieze, A.M., 122
 Frobenius norm, 388, 402
 Furnas, G.W., 414
 Gaber, M.M., 18
 Ganti, V., 122, 265
 Garcia-Molina, H., 18, 190, 226, 265, 382
 Garofalakis, M., 153
 Gaussian elimination, 159
 Gehrke, J., 153, 265
 Generalization, 421
 Generated subgraph, 339
 Genre, 297, 309, 321
 GFS, *see* Google file system
 Ghemawat, S., 67
 Gibbons, P.B., 153, 383
 Gionis, A., 122, 153
 Girvan, M., 382
 Girvan–Newman Algorithm, 333
 Global minimum, 314
 GN Algorithm, *see* Girvan–Newman Algorithm
 Gobioff, H., 67
 Golub, G.H., 414
 Google, 155, 166, 276
 Google file system, 22
 Google+, 326
 Gradient descent, 17, 320, 355, 442
 Granzow, M., 414
 Graph, 42, 54, 325, 326, 361, 368
 Greedy algorithm, 270, 271, 274, 278
 GRGPF Algorithm, 252
 Grouping, 24, 31, 35
 Grouping attribute, 31
 Groupon, 329
 Gruber, R.E., 67
 Guha, S., 266
 Gunda, P.K., 67
 Gyongi, Z., 190
 Hadoop, 22, 67
 Hadoop distributed file system, 22
 Hamming distance, 63, 91, 98
 Harris, M., 321

- Harshman, R., 414
 Hash function, 8, 74, 78, 83, 129, 131, 134
 Hash key, 9, 285
 Hash table, 8, 10, 11, 200, 207, 209, 211, 285, 287, 362
 Haveliwala, T.H., 190
 HDFS, *see* Hadoop distributed file system
 Head, 372
 Heavy hitter, 362
 Henzinger, M., 122
 Hierarchical clustering, 230, 232, 249, 310, 331
 Hinge loss, 441
 HITS, 182
 Hive, 66, 67
 Hopcroft, J.E., 374
 Horn, H., 67
 Howe, B., 67
 Hsieh, W.C., 67
 Hub, 182, 183
 Hyperlink-induced topic search, *see* HITS
 Hyperplane, 436
 Hyracks, 38
 Identical documents, 111
 Identity matrix, 386
 IDF, *see* Inverse document frequency
 Image, 125, 297, 298
 IMDB, *see* Internet Movie Database
 Imielinski, T., 226
 Immediate subset, 218
 Immorlica, N., 122
 Important page, 155
 Impression, 268
 In-component, 159
 Inaccessible page, 178
 Independent rows or columns, 397
 Index, 10, 362
 Indyk, P., 122, 153
 Initialize clusters, 242
 Input, 54
 Insertion, 90
 Instance-based learning, 419
 Interest, 196
 Internet Movie Database, 297, 321
 Interpolation, 450
 Intersection, 31, 33, 38, 71
Into Thin Air, 295
 Inverse document frequency, *see also* TF.IDF, 8
 Inverted index, 155, 268
 Ioannidis, Y.E., 382
 IP packet, 125
 Isard, M., 67
 Isolated component, 160
 Item, 191, 193, 194, 293, 308, 310
 Item profile, 297, 299
 Itemset, 191, 199, 201
 Jaccard distance, 87, 88, 95, 298, 453
 Jaccard similarity, 69, 77, 87, 177
 Jacobsen, H.-A., 67
 Jagadish, H.V., 153
 Jahrer, M., 324
 Jeh, G., 382
 Joachims, T., 458
 Join, *see also* Natural join, *see also* Multiway join, *see also* Star join, 363
 Join task, 40
 K-means, 241
 K-partite graph, 329
 Kahan, W., 414
 Kalyanasundaram, B., 291
 Kamm, D., 321
 Kang, U., 383
 Kannan, R., 414
 Karlin, A., 272
 Kaushik, R., 122
 Kautz, W.H., 153
 Kernel function, 447, 451
 Key component, 129
 Key-value pair, 22–24
 Keyword, 276, 303
 Kleinberg, J.M., 190
 Knuth, D.E., 18
 Koren, Y., 324
 Kosmix, 22
 Krioukov, A., 67
 Kumar, R., 382
 Kumar, R., 18, 67, 190
 Kumar, V., 18
 Label, 326, 416
 Lagrangean multipliers, 48
 Landauer, T.K., 414
 Lang, K.J., 382
 Laplacian matrix, 345
 LCS, *see* Longest common subsequence
 Leaf, 334
 Learning-rate parameter, 424
 Leiser, N., 67
 Length, 137, 367
 Length indexing, 113
 Leskovec, J., 382, 383
 Leung, S.-T., 67
 Likelihood, 351
 Lin, S., 122
 Linden, G., 324
 Linear equations, 159
 Linear separability, 423, 426
 Link, 30, 155, 168
 Link matrix of the Web, 184
 Link spam, 174, 178
 Littlestone, N., 458
 Livny, M., 266

- Local minimum, 314
 Locality, 325, 326
 Locality-sensitive family, 98
 Locality-sensitive function, 94
 Locality-sensitive hashing, 82, 93, 299, 453
 Log likelihood, 356
 Logarithm, 12
 Long tail, 194, 292, 294
 Longest common subsequence, 90
 Lower bound, 57
 LSH, *see* Locality-sensitive hashing
 Machine learning, 2, 17, 303, 415
 Maggioni, M., 414
 Maghoul, F., 18, 190
 Mahalanobis distance, 247
 Mahoney, M.W., 382, 414
 Main memory, 199, 207, 230
 Malewicz, G., 67
 Malik, J., 383
 Manber, U., 122
 Manhattan distance, 88
 Manning, C.P., 18
 Many-many matching, 107
 Many-many relationship, 54, 191
 Many-one matching, 107
 Map task, 22, 23, 25
 Map worker, 25, 26
 Mapping schema, 55
 MapReduce, 15, 19, 22, 28, 168, 169, 217, 261, 363, 370, 434
 Margin, 436
 Market basket, 4, 15, 191, 198
 Markov process, 158, 160, 358
 Martin, G.N., 153
 Master controller, 22, 24, 25
 Master node, 22
 Matching, 273
 Matias, Y., 152
 Matrix, 28, *see also* Transition matrix of the Web, *see also* Stochastic matrix, *see also* Substochastic matrix, 168, 182, *see also* Utility matrix, 312, *see also* Adjacency matrix, *see also* Degree matrix, *see also* Laplacian matrix, *see also* Symmetric matrix
 Matrix multiplication, 35, 36, 59
 Matrix of distances, 395
 Matthew effect, 14
 Maximal itemset, 201
 Maximal matching, 273
 Maximum-likelihood estimation, 351
 McAuley, J., 383
 Mean, *see* Average
 Mechanical Turk, 422
 Median, 136
 Mehta, A., 291
 Melnik, S., 382
 Merging clusters, 233, 237, 247, 251, 255, 259
 Merton, P., 18
 Miller, G.L., 383
 Minhashing, 76, 85, 89, 95, 299
 Miniclust, 245
 Minsky, M., 458
 Minutiae, 107
 Mirrokni, V.S., 122
 Mirror page, 70
 Mitzenmacher, M., 122
 ML, *see* Machine learning
 MLE, *see* Maximum-likelihood estimation
 Model, 351
 Moments, 137
 Monotonicity, 201
 Montavon, G., 458
 Moore–Penrose pseudoinverse, 407
 Most-common elements, 148
 Motwani, R., 122, 153, 226, 265
 Mueller, K.-R., 458
 Multiclass classification, 416, 430
 Multidimensional index, 452
 Multihash Algorithm, 211
 Multiplication, 28, *see also* Matrix multiplication, 168, 182
 Multiset, *see* Bag
 Multistage Algorithm, 209
 Multiway join, 46, 364
 Mumick, I.S., 153
 Mutation, 92
 Name node, *see* Master node
 Natural join, 31, 34, 36, 45
 Naughton, J.F., 67
 Navathe, S.B., 226
 Near-neighbor search, *see* Locality-sensitive hashing
 Nearest neighbor, 17, 419, 447, 455
 Negative border, 218
 Negative example, 423
 Neighbor, 357
 Neighborhood, 367, 375
 Neighborhood profile, 367
 Netflix challenge, 2, 295, 321
 Network, *see* Social network
 Neural net, 419
 Newman, M.E.J., 382
 Newspaper articles, 109, 286, 296
 Non-Euclidean distance, 238, *see also* Cosine distance, *see also* Edit distance, *see also* Hamming distance, *see also* Jaccard distance
 Non-Euclidean space, 252, 254
 Norm, 87, 88
 Normal distribution, 244
 Normalization, 305, 307, 318

- Normalized cut, 344
 NP-complete problem, 338
 Numerical feature, 416, 455
 O'Callaghan, L., 265
 Off-line algorithm, 270
 Olston, C., 67
 Omiecinski, E., 226
 On-line advertising, *see* Advertising
 On-line algorithm, 16, 270
 On-line learning, 421
 On-line retailer, 194, 268, 292, 294
 Open directory, 174, 422
 OR-construction, 95
 Orr, G.B., 458
 Orthogonal vectors, 231, 389
 Orthonormal matrix, 397, 402
 Orthonormal vectors, 389, 393
 Out-component, 159
 Outlier, 230
 Output, 54
 Overfitting, 303, 319, 419, 420, 432, 456
 Overlapping Communities, 350
 Overture, 276
 Own pages, 178
 Paepcke, A., 122
 Page, L., 154, 190
 PageRank, 3, 15, 28, 29, 40, 154, 156, 168
 Pairs, *see* Frequent pairs
 Palmer, C.R., 383
 Pan, J.-Y., 383
 Papert, S., 458
 Parent, 333
 Park, J.S., 226
 Partition, 343
 Pass, 199, 202, 209, 214
 Path, 367
 Paulson, E., 67
 PCA, *see* Principal-component analysis
 PCY Algorithm, 207, 209, 210
 Pearson, K., 414
 Pedersen, J., 190
 Perceptron, 17, 415, 419, 422, 455
 Perfect matching, 273
 Permutation, 76, 81
 PIG, 66
 Pigeonhole principle, 339
 Piotte, M., 324
 Pivotal condensation, 386
 Plagiarism, 69, 195
 Pnuts, 66
 Point, 228, 257
 Point assignment, 230, 241, 332
 Polyzotis, A., 66
 Position indexing, 115, 116
 Positive example, 423
 Positive integer, 147
 Powell, A.L., 265
 Power Iteration, 386
 Power iteration, 387
 Power law, 12
 Predicate, 303
 Prefix indexing, 113, 115, 116
 Pregel, 42
 Principal eigenvector, 158, 385
 Principal-component analysis, 384, 391
 Priority queue, 236
 Priors, 352
 Privacy, 269
 Probe string, 114
 Profile, *see* Item profile, *see* User profile
 Projection, 31, 33
 Pruhs, K.R., 291
 Pseudoinverse, *see* Moore–Penrose
 pseudoinverse
 Puz, N., 67
 Quadratic programming, 442
 Query, 125, 144, 261
 Query example, 447
 R-tree, 265
 Rack, 20
 Radius, 237, 240, 367
 Raghavan, P., 18, 190, 382
 Rahm, E., 382
 Rajagopalan, S., 18, 190, 382
 Ramakrishnan, R., 67, 265, 266
 Ramsey, W., 290
 Random hyperplanes, 99, 299
 Random surfer, 155, 156, 161, 175, 357
 Randomization, 215
 Rank, 397
 Rarest-first order, 287
 Rastogi, R., 153, 266
 Rating, 293, 296
 Reachability, 369
 Recommendation system, 16, 292
 Recursion, 40
 Recursive doubling, 371
 Reduce task, 23, 24
 Reduce worker, 25, 27
 Reducer, 24
 Reducer size, 51, 57
 Reed, B., 67
 Reflexive and transitive closure, 369
 Regression, 416, 451, 455
 Regularization parameter, 441
 Reichsteiner, A., 414
 Reina, C., 265
 Relation, 30
 Relational algebra, 30, 31
 Replication, 22
 Replication rate, 51, 57
 Representation, 253

- Representative point, 250
 Representative sample, 128
 Reservoir sampling, 152
 Restart, 358
 Retained set, 245
 Revenue, 277
 Ripple-carry adder, 147
 RMSE, *see* Root-mean-square error
 Robinson, E., 67
 Rocha, L.M., 414
 Root-mean-square error, 295, 313, 402
 Rosa, M., 382
 Rosenblatt, F., 458
 Rounding data, 307
 Row, *see* Tuple
 Row-orthonormal matrix, 402
 Rowsum, 252
 Royalty, J., 67
 S-curve, 84, 93
 Saberi, A., 291
 Salihoglu, S., 66
 Sample, 215, 218, 221, 223, 242, 249, 253
 Sampling, 127, 141
 Savasere, A., 226
 SCC, *see* Strongly connected component
 Schapire, R.E., 458
 Schema, 30
 Schutze, H., 18
 Score, 105
 Search ad, 268
 Search engine, 166, 181
 Search query, 125, 155, 176, 268, 285
 Second-price auction, 279
 Secondary storage, *see* Disk
 Selection, 31, 33
 Sensor, 124
 Sentiment analysis, 422
 Set, 76, 112, *see also* Itemset
 Set difference, *see* Difference
 Shankar, S., 67
 Shawe-Taylor, J., 458
 Shi, J., 383
 Shim, K., 266
 Shingle, 72, 85, 109
 Shivakumar, N., 226
 Shopping cart, 193
 Shortest paths, 42
 Siddharth, J., 122
 Signature, 75, 78, 85
 Signature matrix, 78, 83
 Silberschatz, A., 153
 Silberstein, A., 67
 Similarity, 4, 15, 69, 191, 299, 306
 Similarity join, 52, 58
 Simrank, 357
 Singleton, R.C., 153
 Singular value, 397, 401, 402
 Singular-value decomposition, 312, 384, 397, 406
 Six degrees of separation, 369
 Sketch, 100
 Skew, 26
 Sliding window, 126, 142, 148, 257
 Smart transitive closure, 372
 Smith, B., 324
 SNAP, 382
 Social Graph, 326
 Social network, 16, 325, 326, 384
 SON Algorithm, 217
 Source, 367
 Space, 87, 228
 Spam, *see also* Term spam, *see also* Link spam, 328, 421
 Spam farm, 178, 180
 Spam mass, 180, 181
 Sparse matrix, 28, 76, 77, 168, 293
 Spectral partitioning, 343
 Spider trap, 161, 164, 184
 Splitting clusters, 255
 SQL, 19, 30, 66
 Squares, 366
 Srikant, R., 226
 Srivastava, U., 67
 Standard deviation, 245, 247
 Standing query, 125
 Stanford Network Analysis Platform, *see* SNAP
 Star join, 50
 Stata, R., 18, 190
 Statistical model, 1
 Status, 287
 Steinbach, M., 18
 Stochastic gradient descent, 320, 445
 Stochastic matrix, 158, 385
 Stop clustering, 234, 238, 240
 Stop words, 7, 74, 110, 194, 298
 Stream, *see* Data stream
 Strength of membership, 355
 String, 112
 Striping, 29, 168, 170
 Strong edge, 328
 Strongly connected component, 159, 374
 Strongly connected graph, 158, 368
 Substochastic matrix, 161
 Suffix length, 116
 Summarization, 3
 Summation, 147
 Sun, J., 414
 Supercomputer, 19
 Superimposed code, *see* Bloom filter, 152
 Supermarket, 193, 214
 Superstep, 43

- Supervised learning, 415, 417
 Support, 192, 216, 218, 221
 Support vector, 437
 Support-vector machine, 17, 415, 419, 436, 455
 Supporting page, 178
 Suri, S., 383
 Surprise number, 137
 SVD, *see* Singular-value decomposition
 SVM, *see* Support-vector machine
 Swami, A., 226
 Symmetric matrix, 346, 384
 Szegedy, M., 152
 Tag, 298, 329
 Tail, 372
 Tail length, 135, 376
 Tan, P.-N., 18
 Target, 367
 Target page, 178
 Tarjan, R.E., 374
 Task, 21
 Taxation, 161, 164, 178, 184
 Taylor expansion, 12
 Taylor, M., 290
 Telephone call, 328
 Teleport set, 175, 176, 181, 358
 Teleportation, 165
 Tendril, 159
 Term, 155
 Term frequency, *see* TF.IDF, 8
 Term spam, 155, 178
 Test set, 420, 426
 TF, *see* Term frequency
 TF.IDF, 7, 8, 298, 418
 Theobald, M., 122
 Thrashing, 170, 207
 Threshold, 84, 150, 192, 216, 221, 422, 428
 TIA, *see* Total Information Awareness
 Timestamp, 143, 258
 Toivonen's Algorithm, 218
 Toivonen, H., 227
 Tomkins, A., 18, 67, 190, 382
 Tong, H., 383
 Topic-sensitive PageRank, 174, 181
 Toscher, A., 324
 Total Information Awareness, 5
Touching the Void, 295
 Training example, 416
 Training rate, 426
 Training set, 415, 416, 422, 431
 Transaction, *see* Basket
 Transition matrix, 358
 Transition matrix of the Web, 157, 168, 171, 384
 Transitive closure, 40, 369
 Transitive reduction, 374
 Transpose, 184
 Transposition, 92
 Tree, 236, 252, 253, *see also* Decision tree
 Triangle, 361
 Triangle inequality, 87
 Triangular matrix, 200, 209
 Tripartite graph, 330
 Triples method, 200, 209
 TrustRank, 180, 181
 Trustworthy page, 181
 Tsourakakis, C.E., 383
 Tube, 160
 Tuple, 30
 Tuzhilin, A., 324
 Twitter, 16, 286, 326
 Ullman, J.D., 18, 66, 67, 226, 265, 382
 Undirected graph, *see* Graph
 Union, 31, 33, 38, 71
 Unit vector, 385, 389
 Universal set, 112
 Unsupervised learning, 415
 User, 293, 308, 310
 User profile, 301
 Utility matrix, 293, 296, 312, 384
 UV-decomposition, 312, 321, 384, 446
 VA file, 452
 Valduriez, P., 383
 Van Loan, C.F., 414
 Vapnik, V.N., 458
 Variable, 138
 Vassilvitskii, S., 383
 Vazirani, U., 291
 Vazirani, V., 291
 Vector, 28, 87, 91, 157, 168, 182, 183, 228
 Vigna, S., 382
 Vitter, J., 153
 Volume (of a set of nodes), 344
 von Ahn, L., 299, 324
 von Luxburg, U., 383
 Voronoi diagram, 448
 Wall, M.E., 414
 Wall-clock time, 46
 Wallach, D.A., 67
 Wang, J., 321
 Wang, W., 122
 Weak edge, 328
 Weaver, D., 67
 Web structure, 159
 Weight, 422
 Weiner, J., 18, 190
 Whizbang Labs, 2
 Widom, J., 18, 67, 153, 265, 382
 Wikipedia, 328, 422
 Window, *see* Sliding window, *see* Decaying window
 Windows, 11

Winnow Algorithm, 427
Word, 194, 229, 298
Word count, 23
Worker process, 25
Workflow, 39, 40, 44
Working store, 124
Xiao, C., 122
Xie, Y., 414
Yahoo, 276, 299
Yang, J., 383
Yerneni, R., 67
York, J., 324
Yu, J.X., 122
Yu, P.S., 226
Yu, Y., 67
Zhang, H., 414
Zhang, T., 266
Zipf's law, 14, *see* Power law
Zoeter, O., 290