

Index

- A-Priori Algorithm, 212, 213, 219
- Accessible page, 187
- Active learning, 446
- Ad-hoc query, 134
- Adjacency matrix, 363
- Adomavicius, G., 340
- Advertising, 16, 116, 204, 281
- Adwords, 290
- Affiliation-Graph model, 371
- Afrati, F.N., 70, 402
- Agglomerative clustering, *see* Hierarchical clustering
- Aggregation, 34, 37
- Agrawal, R., 238
- Alon, N., 162
- Alon-Matias-Szegedy Algorithm, 146
- Amplification, 101
- Analytic query, 53
- AND-construction, 101
- Anderson, C., 340, 341
- Andoni, A., 129
- ANF, *see* Approximate neighborhood function
- ANF Algorithm, 396
- Apache, 24, 71
- Approximate neighborhood function, 396
- Arc, 386
- Archive, 132
- Ask, 192
- Association rule, 205, 207
- Associativity, 27
- Attribute, 33
- Auction, 293
- Austern, M.H., 71
- Authority, 192
- Average, 144
- B-tree, 280
- Babcock, B., 162, 280
- Babu, S., 162
- Backstrom, L., 402
- Bag, 40, 76
- Balance Algorithm, 293
- Balazinska, M., 70
- Band, 88
- Bandwidth, 22
- Basket, *see* Market basket, 202, 204, 205, 234
- Batch gradient descent, 471
- Batch learning, 445
- Bayes net, 4
- BDMO Algorithm, 271
- Beer and diapers, 206
- Bell, R., 341
- Bellkor's Pragmatic Chaos, 310
- Berkhin, P., 200
- Berrar, D.P., 437
- Betweenness, 351
- BFR Algorithm, 254, 257
- BFS, *see* Breadth-first search
- Bi-clique, 357
- Bid, 291, 293, 300, 301
- BigTable, 70
- Bik, A.J.C., 71
- Binary Classification, 440
- Biomarker, 205
- Bipartite graph, 287, 347, 357, 358
- BIRCH Algorithm, 280
- Birrell, A., 71
- Bitmap, 220, 221
- Block, 11, 21, 180
- Blog, 188
- Bloom filter, 140, 218

- Bloom, B.H., 162
- Blum, A., 483
- Bohannon, P., 70
- Boldi, P., 402
- Bonferroni correction, 5
- Bonferroni's principle, 4, 5
- Bookmark, 186
- Boral, H., 403
- Borkar, V., 70
- Bottou, L., 483
- Bradley, P.S., 280
- Breadth-first search, 351
- Brick-and-mortar retailer, 204, 308, 309
- Brin, S., 200
- Broad matching, 293
- Broder, A.Z., 18, 129, 200
- Bu, Y., 70
- Bucket, 9, 137, 152, 156, 218, 271
- Budget, 292, 299
- Budiu, M., 71
- Burges, C.J.C., 483
- Burrows, M., 70

- Candidate itemset, 215, 228
- Candidate pair, 88, 219, 222
- Carey, M., 70
- Categorical feature, 440, 480
- Centroid, 243, 246, 252, 255, 259
- Chabbert, M., 341
- Chandra, T., 70
- Chang, F., 70
- Characteristic matrix, 81
- Charikar, M.S., 129
- Chaudhuri, S., 129
- Checkpoint, 46
- Chen, M.-S., 238
- Child, 351
- Cholera, 3
- Chronicle data model, 161
- Chunk, 24, 228, 258
- CineMatch, 337
- Classifier, 318, 439
- Click stream, 133
- Click-through rate, 285, 291
- Clique, 357
- Cloud computing, 15
- CloudStore, 24
- Cluster computing, 21, 22
- Cluster tree, 266, 267
- Clustera, 41, 69
- Clustering, 3, 16, 241, 325, 343, 349, 439
- Clustroid, 246, 252
- Collaboration network, 346
- Collaborative filtering, 4, 17, 75, 281, 307, 321, 347
- Column-orthonormal matrix, 419
- Combiner, 27, 177, 179
- Communication cost, 22, 47, 384
- Community, 17, 343, 354, 357, 381
- Community-affiliation graph, 371
- Commutativity, 27
- Competitive ratio, 16, 286, 289, 294
- Complete graph, 357, 358
- Compressed set, 258
- Compute node, 21, 22
- Computer game, 315
- Computing cloud, *see* Cloud computing
- Concept, 420
- Concept space, 425
- Confidence, 206, 207
- Content-based recommendation, 307, 312
- Convergence, 451
- Cooper, B.F., 70
- Coordinates, 242
- Cortes, C., 484
- Cosine distance, 95, 105, 313, 318, 426
- Counting ones, 150, 271
- Covering an output, 59
- Craig's List, 282
- Craswell, N., 305
- Credit, 352
- Cristianini, N., 483
- Cross-Validation, 445
- Crowdsourcing, 446
- CUR-decomposition, 405, 428
- CURE Algorithm, 262, 266
- Currey, J., 71

- Curse of dimensionality, 244, 268, 478, 481
- Cut, 362
- Cyclic permutation, 87
- Cylinder, 12
- Czajkowski, G., 71
- DAG, *see* Directed acyclic graph
- Darts, 140
- Das Sarma, A., 70
- Dasgupta, A., 403
- Data mining, 1
- Data stream, 16, 232, 270, 284, 460
- Data-stream-management system, 132
- Database, 16
- Datar, M., 129, 162, 280
- Datar-Gionis-Indyk-Motwani Algorithm, 151
- Dead end, 167, 170, 171, 193
- Dean, J., 70, 71
- Decaying window, 157, 234
- Decision tree, 17, 318, 443, 444, 481
- Deerwester, S., 436
- Degree, 359, 381
- Degree matrix, 363
- Dehnert, J.C., 71
- del.icio.us, 314, 347
- Deletion, 95
- Dense matrix, 31, 428
- Density, 251, 253
- Depth-first search, 393
- Determinant, 407
- DeWitt, D.J., 71
- DFS, *see* Distributed file system
- Diagonal matrix, 419
- Diameter, 251, 253, 388
- Diapers and beer, 204
- Difference, 33, 36, 40
- Dimension table, 53
- Dimensionality reduction, 17, 328, 405, 478
- Directed acyclic graph, 351
- Directed graph, 386
- Discard set, 258
- Disk, 11, 209, 243, 266
- Disk block, *see* Block
- Display ad, 282, 283
- Distance measure, 92, 241, 349
- Distinct elements, 142, 145
- Distributed file system, 21, 23, 202, 209
- DMOZ, *see* Open directory
- Document, 74, 77, 205, 242, 301, 313, 314, 442
- Document frequency, *see* Inverse document frequency
- Domain, 190
- Dot product, 95
- Drineas, P., 436
- Dryad, 69
- DryadLINQ, 70
- Dual construction, 348
- Dubitzky, W., 437
- Dumais, S.T., 436
- Dup-elim task, 43
- e , 12
- Edit distance, 95, 98
- Eigenpair, 406
- Eigenvalue, 167, 364, 405, 416, 417
- Eigenvector, 167, 364, 405, 411, 416
- Email, 346
- Energy, 424
- Ensemble, 319, 481
- Entity resolution, 110
- Equijoin, 34
- Erlingsson, I., 71
- Ernst, M., 70
- Ethernet, 21, 22
- Euclidean distance, 93, 107, 477
- Euclidean space, 93, 97, 242, 243, 246, 262
- Exponentially decaying window, *see* Decaying window
- Extrapolation, 476
- Facebook, 17, 186, 344
- Fact table, 53
- Failure, 23, 29, 42–44
- Faloutsos, C., 403, 437

- False negative, 88, 99, 227
- False positive, 88, 99, 140, 227
- Family of functions, 100
- Fang, M., 238
- Fayyad, U.M., 280
- Feature, 266, 312–314
- Feature selection, 446
- Feature vector, 440, 480
- Fetterly, D., 71
- Fikes, A., 70
- File, 23, 24, 209, 227
- Filtering, 139
- Fingerprint, 113
- First-price auction, 293
- Fixedpoint, 102, 192
- Flajolet, P., 162
- Flajolet-Martin Algorithm, 143, 395
- Flow graph, 41
- Fortunato, S., 402
- Fotakis, D., 402
- French, J.C., 280
- Frequent bucket, 219, 221
- Frequent itemset, 4, 202, 212, 214, 358, 439
- Frequent pairs, 213
- Frequent-items table, 214
- Freund, Y., 484
- Friends, 344
- Friends relation, 52
- Frieze, A.M., 129
- Frobenius norm, 409, 423
- Furnas, G.W., 436
- Gaber, M.M., 18
- Ganti, V., 129, 280
- Garcia-Molina, H., 18, 200, 238, 280, 403
- Garofalakis, M., 162
- Gaussian elimination, 168
- Gehrke, J., 162, 280
- Generalization, 445
- Generated subgraph, 357
- Genre, 312, 324, 338
- GFS, *see* Google file system
- Ghemawat, S., 70, 71
- Gibbons, P.B., 162, 403
- Gionis, A., 129, 162
- Girvan, M., 403
- Girvan-Newman Algorithm, 351
- Global minimum, 330
- GN Algorithm, *see* Girvan-Newman Algorithm
- Gobioff, H., 71
- Golub, G.H., 436
- Google, 164, 175, 290
- Google file system, 24
- Google+, 344
- Gradient descent, 17, 336, 373, 467
- Granzow, M., 437
- Graph, 45, 57, 343, 344, 380, 387
- Greedy algorithm, 284, 285, 288, 292
- GRGPF Algorithm, 266
- Grouping, 26, 34, 37
- Grouping attribute, 34
- Groupon, 347
- Gruber, R.E., 70
- Guha, S., 280
- Gunda, P.K., 71
- Gyongyi, Z., 200
- Hadoop, 24, 71
- Hadoop distributed file system, 24
- Hamming distance, 67, 96, 104
- Harris, M., 338
- Harshman, R., 436
- Hash function, 9, 79, 83, 88, 137, 140, 143
- Hash key, 9, 300
- Hash table, 9, 10, 12, 211, 218, 221, 222, 300, 302, 381
- Haveliwala, T.H., 200
- HDFS, *see* Hadoop distributed file system
- Head, 392
- Heavy hitter, 381
- Henzinger, M., 129
- Hierarchical clustering, 243, 245, 263, 326, 349
- Hinge loss, 466
- HITS, 192

- Hive, 70, 71
- Hopcroft, J.E., 393
- Horn, H., 71
- Howe, B., 70
- Hsieh, W.C., 70
- Hub, 192
- Hyperlink-induced topic search, *see* HITS
- Hyperplane, 461
- Hyracks, 41
- Identical documents, 118
- Identity matrix, 407
- IDF, *see* Inverse document frequency
- Image, 133, 313, 314
- IMDB, *see* Internet Movie Database
- Imielinski, T., 238
- Immediate subset, 230
- Immorlica, N., 129
- Important page, 164
- Impression, 282
- In-component, 169
- Inaccessible page, 187
- Independent rows or columns, 419
- Index, 10, 381
- Indyk, P., 129, 162
- Initialize clusters, 255
- Input, 57
- Insertion, 95
- Instance-based learning, 443
- Interest, 206
- Internet Movie Database, 312, 338
- Interpolation, 476
- Intersection, 33, 36, 40, 77
- Into Thin Air*, 311
- Inverse document frequency, *see* TF.IDF
- 8
- Inverted index, 164, 282
- Ioannidis, Y.E., 403
- IP packet, 133
- Isard, M., 71
- Isolated component, 170
- Item, 202, 204, 205, 308, 324, 325
- Item profile, 312, 315
- Itemset, 202, 210, 212
- Jaccard distance, 92, 94, 100, 313, 479
- Jaccard similarity, 74, 82, 92, 187
- Jacobsen, H.-A., 70
- Jagadish, H.V., 162
- Jahrer, M., 341
- Jeh, G., 403
- Joachims, T., 484
- Join, *see* Natural join, *see* Multiway join, *see* Star join, 383
- Join task, 43
- K-means, 254
- K-partite graph, 347
- Kahan, W., 436
- Kalyanasundaram, B., 306
- Kamm, D., 338
- Kang, U., 403
- Kannan, R., 436
- Karlin, A., 286
- Kaushik, R., 129
- Kautz, W.H., 162
- Kernel function, 473, 477
- Key component, 137
- Key-value pair, 25–27
- Keyword, 291, 319
- Kleinberg, J.M., 200
- Knuth, D.E., 18
- Koren, Y., 341
- Kosmix, 24
- Krioukov, A., 71
- Kumar, R., 18, 71, 200, 403
- Kumar, V., 18
- Label, 344, 440
- Lagrangean multipliers, 51
- Landauer, T.K., 436
- Lang, K.J., 403
- Laplacian matrix, 364
- LCS, *see* Longest common subsequence
- Leaf, 352
- Learning-rate parameter, 448
- Leiser, N., 71
- Length, 146, 387
- Length indexing, 119
- Leskovec, J., 402–404
- Leung, S.-T., 71

- Likelihood, 369
- Lin, S., 129
- Linden, G., 341
- Linear equations, 168
- Linear separability, 447, 451
- Link, 33, 164, 178
- Link matrix of the Web, 193
- Link spam, 183, 187
- Littlestone, N., 484
- Livny, M., 280
- Local minimum, 330
- Locality, 344
- Locality-sensitive family, 104
- Locality-sensitive function, 99
- Locality-sensitive hashing, 88, 99, 314, 479
- Log likelihood, 374
- Logarithm, 12
- Long tail, 204, 308, 309
- Longest common subsequence, 96
- Lower bound, 61
- LSH, *see* Locality-sensitive hashing

- Machine learning, 2, 17, 318, 439
- Maggioni, M., 436
- Maghoul, F., 18, 200
- Mahalanobis distance, 261
- Mahoney, M.W., 403, 436
- Main memory, 209, 210, 218, 243
- Malewicz, G, 71
- Malik, J., 403
- Manber, U., 129
- Manhattan distance, 93
- Manning, C.P., 18
- Many-many matching, 113
- Many-many relationship, 57, 202
- Many-one matching, 113
- Map task, 25, 27
- Map worker, 28, 29
- Mapping schema, 58
- MapReduce, 15, 21, 24, 30, 177, 179, 229, 275, 383, 390, 458
- Margin, 461
- Market basket, 4, 16, 201, 202, 209
- Markov process, 167, 170, 377
- Martin, G.N., 162
- Master controller, 25, 26, 28
- Master node, 24
- Matching, 287
- Matias, Y., 162
- Matrix, 31, *see* Transition matrix of the Web, *see* Stochastic matrix, *see* Substochastic matrix, 177, 192, *see* Utility matrix, 328, *see* Adjacency matrix, *see* Degree matrix, *see* Laplacian matrix, *see* Symmetric matrix
- Matrix multiplication, 38, 39, 62
- Matrix of distances, 417
- Matthew effect, 14
- Maximal itemset, 212
- Maximal matching, 287
- Maximum-likelihood estimation, 369
- McAuley, J., 404
- Mean, *see* Average
- Mechanical Turk, 446
- Median, 144
- Mehta, A., 306
- Melnik, S., 403
- Merging clusters, 246, 249, 260, 264, 269, 273
- Merton, P., 18
- Miller, G.L., 403
- Minhashing, 81, 91, 94, 101, 314
- Miniclustet, 258
- Minsky, M., 484
- Minutiae, 113
- Mirrokn, V.S., 129
- Mirror page, 75
- Mitzenmacher, M., 129
- ML, *see* Machine learning
- MLE, *see* Maximum-likelihood estimation
- Model, 369
- Moments, 145
- Monotonicity, 212
- Montavon, G., 483
- Moore-Penrose pseudoinverse, 429
- Most-common elements, 157
- Motwani, R., 129, 162, 238, 280

- Mueller, K.-R., 483
- Multiclass classification, 440, 455
- Multidimensional index, 478
- Multihash Algorithm, 222
- Multiplication, 31, *see* Matrix multiplication, 177, 192
- Multiset, *see* Bag
- Multistage Algorithm, 220
- Multiway join, 49, 383
- Mumick, I.S., 162
- Mutation, 98

- Name node, *see* Master node
- Natural join, 34, 37, 38, 48
- Naughton, J.F., 71
- Navathe, S.B., 238
- Near-neighbor search, *see* Locality-sensitive hashing
- Nearest neighbor, 17, 444, 472, 481
- Negative border, 230
- Negative example, 447
- Neighbor, 376
- Neighborhood, 387, 395
- Neighborhood profile, 387
- Netflix challenge, 2, 310, 337
- Network, *see* Social network
- Neural net, 443
- Newman, M.E.J., 403
- Newspaper articles, 115, 301, 310
- Non-Euclidean distance, 252, *see* Cosine distance, *see* Edit distance, *see* Hamming distance, *see* Jac-card distance
- Non-Euclidean space, 266, 268
- Norm, 93
- Normal distribution, 257
- Normalization, 321, 323, 334
- Normalized cut, 363
- NP-complete problem, 357
- Numerical feature, 440, 480

- O’Callaghan, L., 280
- Off-line algorithm, 284
- Olston, C., 71
- Omiecinski, E., 238

- On-line advertising, *see* Advertising
- On-line algorithm, 16, 284
- On-line learning, 445
- On-line retailer, 204, 282, 308, 309
- Open directory, 184, 446
- OR-construction, 101
- Orr, G.B., 483
- Orthogonal vectors, 244, 410
- Orthonormal matrix, 419, 424
- Orthonormal vectors, 411, 414
- Out-component, 169
- Outlier, 243
- Output, 57
- Overfitting, 319, 336, 443, 444, 457, 481
- Overlapping Communities, 369
- Overture, 291
- Own pages, 188

- Paepcke, A., 129
- Page, L., 163, 200
- PageRank, 3, 16, 31, 32, 42, 163, 165, 177
- Pairs, *see* Frequent pairs
- Palmer, C.R., 403
- Pan, J.-Y., 403
- Papert, S., 484
- Parent, 351
- Park, J.S., 238
- Partition, 361
- Pass, 210, 213, 221, 226
- Path, 387
- Paulson, E., 71
- PCA, *see* Principal-component analysis
- PCY Algorithm, 218, 221, 222
- Pedersen, J., 200
- Perceptron, 17, 439, 443, 447, 481
- Perfect matching, 287
- Permutation, 82, 87
- PIG, 70
- Pigeonhole principle, 357
- Piotte, M., 341
- Pivotal condensation, 407
- Plagiarism, 75, 205

- Pnuts, 70
- Point, 241, 271
- Point assignment, 243, 254, 350
- Polyzotis, A., 70
- Position indexing, 121, 122
- Positive example, 447
- Positive integer, 156
- Powell, A.L., 280
- Power Iteration, 407
- Power iteration, 408
- Power law, 13
- Predicate, 318
- Prefix indexing, 119, 121, 122
- Pregel, 45
- Principal eigenvector, 167, 407
- Principal-component analysis, 405, 412
- Priority queue, 249
- Priors, 371
- Privacy, 284
- Probe string, 121
- Profile, *see* Item profile, *see* User profile
- Projection, 33, 36
- Pruhs, K.R., 306
- Pseudoinverse, *see* Moore-Penrose pseudoinverse
- Puz, N., 70
- Quadratic programming, 467
- Query, 134, 153, 275
- Query example, 473
- R-tree, 280
- Rack, 22
- Radius, 251, 253, 387
- Raghavan, P., 18, 200, 403
- Rahm, E., 403
- Rajagopalan, S., 18, 200, 403
- Ramakrishnan, R., 70, 280
- Ramsey, W., 305
- Random hyperplanes, 105, 314
- Random surfer, 164, 165, 170, 184, 376
- Randomization, 226
- Rank, 418
- Rarest-first order, 301
- Rastogi, R., 162, 280
- Rating, 308, 311
- Reachability, 389
- Recommendation system, 16, 307
- Recursion, 42
- Recursive doubling, 391
- Reduce task, 25, 27
- Reduce worker, 28, 30
- Reducer, 27
- Reducer size, 54, 60
- Reed, B., 71
- Reflexive and transitive closure, 389
- Regression, 440, 477, 481
- Regularization parameter, 466
- Reichsteiner, A., 437
- Reina, C., 280
- Relation, 33
- Relational algebra, 32, 33
- Replication, 24
- Replication rate, 54, 61
- Representation, 266
- Representative point, 263
- Representative sample, 137
- Reservoir sampling, 162
- Restart, 377
- Retained set, 258
- Revenue, 292
- Ripple-carry adder, 156
- RMSE, *see* Root-mean-square error
- Robinson, E., 71
- Rocha, L.M., 437
- Root-mean-square error, 310, 329, 423
- Rosa, M., 402
- Rosenblatt, F., 484
- Rounding data, 323
- Row, *see* Tuple
- Row-orthonormal matrix, 424
- Rowsum, 266
- Royalty, J., 71
- S-curve, 89, 99
- Saberi, A., 306
- Salihoglu, S., 70
- Sample, 226, 230, 233, 235, 255, 263, 267

- Sampling, 136, 150
- Savasere, A., 238
- SCC, *see* Strongly connected component
- Schapire, R.E., 484
- Schema, 33
- Schutze, H., 18
- Score, 111
- Search ad, 282
- Search engine, 175, 191
- Search query, 133, 164, 186, 282, 300
- Second-price auction, 293
- Secondary storage, *see* Disk
- Selection, 33, 35
- Sensor, 133
- Sentiment analysis, 447
- Set, 81, 118, *see* Itemset
- Set difference, *see* Difference
- Shankar, S., 71
- Shawe-Taylor, J., 483
- Shi, J., 403
- Shim, K., 280
- Shingle, 77, 91, 116
- Shivakumar, N., 238
- Shopping cart, 204
- Shortest paths, 45
- Siddharth, J., 129
- Signature, 80, 83, 91
- Signature matrix, 83, 88
- Silberschatz, A., 162
- Silberstein, A., 70
- Similarity, 4, 15, 74, 201, 314, 322
- Similarity join, 55, 61
- Simrank, 376
- Singleton, R.C., 162
- Singular value, 419, 423, 424
- Singular-value decomposition, 328, 405, 418, 428
- Six degrees of separation, 389
- Sketch, 106
- Skew, 28
- Sliding window, 134, 150, 157, 271
- Smart transitive closure, 392
- Smith, B., 341
- SNAP, 402
- Social Graph, 344
- Social network, 17, 343, 344, 405
- SON Algorithm, 228
- Source, 386
- Space, 92, 93, 241
- Spam, *see* Term spam, *see* Link spam, 346, 445
- Spam farm, 187, 190
- Spam mass, 190, 191
- Sparse matrix, 31, 81, 83, 177, 178, 308
- Spectral partitioning, 361
- Spider trap, 170, 173, 193
- Splitting clusters, 269
- SQL, 22, 33, 70
- Squares, 385
- Srikant, R., 238
- Srivastava, U., 70, 71
- Standard deviation, 259, 261
- Standing query, 134
- Stanford Network Analysis Platform, *see* SNAP
- Star join, 53
- Stata, R., 18, 200
- Statistical model, 1
- Status, 301
- Steinbach, M., 18
- Stochastic gradient descent, 336, 471
- Stochastic matrix, 167, 407
- Stop clustering, 247, 251, 253
- Stop words, 8, 79, 116, 205, 313
- Stream, *see* Data stream
- Strength of membership, 374
- String, 118
- Striping, 32, 177, 179
- Strong edge, 346
- Strongly connected component, 169, 393
- Strongly connected graph, 167, 388
- Substochastic matrix, 170
- Suffix length, 123
- Summarization, 3
- Summation, 156
- Sun, J., 437
- Supercomputer, 21
- Superimposed code, *see* Bloom filter, 161

- Supermarket, 204, 226
- Superstep, 46
- Supervised learning, 439, 441
- Support, 202, 227, 228, 230, 232
- Support vector, 462
- Support-vector machine, 17, 439, 444, 461, 481
- Supporting page, 188
- Suri, S., 403
- Surprise number, 146
- SVD, *see* Singular-value decomposition
- SVM, *see* Support-vector machine
- Swami, A., 238
- Symmetric matrix, 365, 406
- Szegedy, M., 162

- Tag, 314, 347
- Tail, 392
- Tail length, 143, 395
- Tan, P.-N., 18
- Target, 386
- Target page, 188
- Tarjan, R.E., 393
- Task, 23
- Taxation, 170, 173, 188, 193
- Taylor expansion, 12
- Taylor, M., 305
- Telephone call, 346
- Teleport set, 184, 185, 190, 377
- Teleportation, 174
- Tendril, 169
- Term, 164
- Term frequency, *see* TF.IDF, 8
- Term spam, 164, 187
- Test set, 444, 451
- TF, *see* Term frequency
- TF.IDF, 7, 8, 313, 443
- Theobald, M., 129
- Thrashing, 179, 218
- Threshold, 89, 159, 202, 228, 232, 447, 453
- TIA, *see* Total Information Awareness
- Timestamp, 151, 272
- Toivonen's Algorithm, 230
- Toivonen, H., 238
- Tomkins, A., 18, 71, 200, 403
- Tong, H., 403
- Topic-sensitive PageRank, 183, 190
- Toscher, A., 341
- Total Information Awareness, 5
- Touching the Void*, 311
- Training example, 440
- Training rate, 451
- Training set, 439, 440, 446, 456
- Transaction, *see* Basket
- Transition matrix, 377
- Transition matrix of the Web, 166, 177, 178, 180, 405
- Transitive closure, 43, 389
- Transitive reduction, 393
- Transpose, 193
- Transposition, 98
- Tree, 248, 266, 267, *see* Decision tree
- Triangle, 380
- Triangle inequality, 93
- Triangular matrix, 211, 220
- Tripartite graph, 347
- Triples method, 211, 220
- TrustRank, 190
- Trustworthy page, 190
- Tsourakakis, C.E., 403
- Tube, 170
- Tuple, 33
- Tuzhilin, A., 340
- Twitter, 17, 301, 344

- Ullman, J.D., 18, 70, 71, 238, 280, 402
- Undirected graph, *see* Graph
- Union, 33, 36, 40, 77
- Unit vector, 406, 411
- Universal set, 118
- Unsupervised learning, 439
- User, 308, 324, 325
- User profile, 316
- Utility matrix, 308, 311, 328, 405
- UV-decomposition, 328, 338, 405, 471

- VA file, 478
- Valduriez, P., 403
- Van Loan, C.F., 436

- Vapnik, V.N., 484
Variable, 146
Vassilvitskii, S., 403
Vazirani, U., 306
Vazirani, V., 306
Vector, 31, 93, 97, 167, 177, 192, 193, 242
Vigna, S., 402
Vitter, J., 162
Volume (of a set of nodes), 363
von Ahn, L., 315, 341
von Luxburg, U., 403
Voronoi diagram, 473
- Wall, M.E., 437
Wall-clock time, 49
Wallach, D.A., 70
Wang, J., 338
Wang, W., 129
Weak edge, 346
Weaver, D., 70
Web structure, 169
Weight, 447
Weiner, J., 18, 200
Whizbang Labs, 2
Widom, J., 18, 71, 162, 280, 403
Wikipedia, 346, 446
Window, *see* Sliding window, *see* Decaying window
Windows, 12
Winnnow Algorithm, 451
Word, 205, 242, 313
Word count, 26
Worker process, 28
Workflow, 41, 43, 47
Working store, 132
- Xiao, C., 129
Xie, Y., 437
- Yahoo, 291, 314
Yang, J., 403, 404
Yerneni, R., 70
York, J., 341
Yu, J.X., 129
Yu, P.S., 238
Yu, Y., 71
Zhang, H., 437
Zhang, T., 280
Zipf's law, 15, *see* Power law
Zoeter, O., 305