Index

A-Priori Algorithm, 210, 211, 217
Accessible page, 185
Active learning, 444
Ad-hoc query, 132
Adjacency matrix, 361
Adomavicius, G., 338
Advertising, 16, 114, 202, 279
Adwords, 288
Affiliation-Graph model, 369
Afrati, F.N., 68, 400
Agglomerative clustering, *see* Hierarchical clustering
Aggregation, 32, 35
Agrawal, R., 236
Alon, N., 160
Alon-Matias-Szegedy Algorithm, 144
Amplification, 99
Analytic query, 51
AND-construction, 99
Anderson, C., 338, 339
Andoni, A., 127
ANF, *see* Approximate neighborhood function
ANF Algorithm, 394
Apache, 22, 69
Approximate neighborhood function, 394
Arc, 384
Archive, 130
Ask, 190
Association rule, 203, 205
Associativity, 25
Attribute, 31
Auction, 291
Austern, M.H., 69
Authority, 190
Average, 142
B-tree, 278
Babcock, B., 160, 278
Babu, S., 160
Backstrom, L., 400
Bag, 38, 74
Balance Algorithm, 291
Balazinska, M., 68
Band, 86
Bandwidth, 20
Basket, *see* Market basket, 200, 202, 203, 232
Batch gradient descent, 469
Batch learning, 443
Bayes net, 4
BDMO Algorithm, 269
Beer and diapers, 204
Bell, R., 339
Bellkor's Pragmatic Chaos, 308
Berkhin, P., 198
Berrar, D.P., 435
Betweenness, 349
BFR Algorithm, 252, 255
BFS, *see* Breadth-first search
Bi-clique, 355
Bid, 289, 291, 298, 299
BigTable, 68
Bik, A.J.C., 69
Binary Classification, 438
Biomarker, 203
Bipartite graph, 285, 345, 355, 356
BIRCH Algorithm, 278
Birrell, A., 69
Bitmap, 218, 219
Block, 11, 19, 178
Blog, 186
Bloom filter, 138, 216

Bloom, B.H., 160
Blum, A., 481
Bohannon, P., 68
Boldi, P., 400
Bonferroni correction, 5
Bonferroni's principle, 4, 5
Bookmark, 184
Boral, H., 401
Borkar, V., 68
Bottou, L., 481
Bradley, P.S., 278
Breadth-first search, 349
Brick-and-mortar retailer, 202, 306, 307
Brin, S., 198
Broad matching, 291
Broder, A.Z., 18, 127, 198
Bu, Y., 68
Bucket, 9, 135, 150, 154, 216, 269
Budget, 290, 297
Budiu, M., 69
Burges, C.J.C., 481
Burrows, M., 68

Candidate itemset, 213, 226
Candidate pair, 86, 217, 220
Carey, M., 68
Categorical feature, 438, 478
Centroid, 241, 244, 250, 253, 257
Chabbert, M., 339
Chandra, T., 68
Chang, F., 68
Characteristic matrix, 79
Charikar, M.S., 127
Chaudhuri, S., 127
Checkpoint, 44
Chen, M.-S., 236
Child, 349
Cholera, 3
Chronicle data model, 159
Chunk, 22, 226, 256
CineMatch, 335
Classifier, 316, 437
Click stream, 131
Click-through rate, 283, 289
Clique, 355
Cloud computing, 15
CloudStore, 22
Cluster computing, 19, 20
Cluster tree, 264, 265
Clustera, 39, 67
Clustering, 3, 16, 239, 323, 341, 347, 437
Clustroid, 244, 250
Collaboration network, 344
Collaborative filtering, 4, 17, 73, 279, 305, 319, 345
Column-orthonormal matrix, 417
Combiner, 25, 175, 177
Communication cost, 20, 45, 382
Community, 341, 352, 355, 379
Community-affiliation graph, 369
Commutativity, 25
Competitive ratio, 16, 284, 287, 292
Complete graph, 355, 356
Compressed set, 256
Compute node, 19, 20
Computer game, 313
Computing cloud, *see* Cloud computing
Concept, 418
Concept space, 423
Confidence, 204, 205
Content-based recommendation, 305, 310
Convergence, 449
Cooper, B.F., 68
Coordinates, 240
Cortes, C., 482
Cosine distance, 93, 103, 311, 316, 424
Counting ones, 148, 269
Covering an output, 57
Craig's List, 280
Craswell, N., 303
Credit, 350
Cristianini, N., 481
Cross-Validation, 443
Crowdsourcing, 444
CUR-decomposition, 403, 426
CURE Algorithm, 260, 264
Currey, J., 69

Curse of dimensionality, 242, 266, 476, 479
Cut, 360
Cyclic permutation, 85
Cylinder, 12
Czajkowski, G., 69

DAG, *see* Directed acyclic graph
Darts, 138
Das Sarma, A., 68
Dasgupta, A., 401
Data mining, 1
Data stream, 16, 230, 268, 282, 458
Data-stream-management system, 130
Database, 16
Datar, M., 127, 160, 278
Datar-Gionis-Indyk-Motwani Algorithm, 149
Dead end, 165, 168, 169, 191
Dean, J., 68, 69
Decaying window, 155, 232
Decision tree, 316, 441, 442, 479
Deerwester, S., 434
Degree, 357, 379
Degree matrix, 361
Dehnert, J.C., 69
del.icio.us, 312, 345
Deletion, 93
Dense matrix, 29, 426
Density, 249, 251
Depth-first search, 391
Determinant, 405
DeWitt, D.J., 69
DFS, *see* Distributed file system
Diagonal matrix, 417
Diameter, 249, 251, 386
Diapers and beer, 202
Difference, 31, 34, 38
Dimension table, 51
Dimensionality reduction, 326, 403, 476
Directed acyclic graph, 349
Directed graph, 384
Discard set, 256
Disk, 11, 207, 241, 264
Disk block, *see* Block
Display ad, 280, 281
Distance measure, 90, 239, 347
Distinct elements, 140, 143
Distributed file system, 19, 21, 200, 207
DMOZ, *see* Open directory
Document, 72, 75, 203, 240, 299, 311, 312, 440
Document frequency, *see* Inverse document frequency
Domain, 188
Dot product, 93
Drineas, P., 434
Dryad, 67
DryadLINQ, 68
Dual construction, 346
Dubitzky, W., 435
Dumais, S.T., 434
Dup-elim task, 41

e, 12
Edit distance, 93, 96
Eigenpair, 404
Eigenvalue, 165, 362, 403, 414, 415
Eigenvector, 165, 362, 403, 409, 414
Email, 344
Energy, 422
Ensemble, 317, 479
Entity resolution, 108
Equijoin, 32
Erlingsson, I., 69
Ernst, M., 68
Ethernet, 19, 20
Euclidean distance, 91, 105, 475
Euclidean space, 91, 95, 240, 241, 244, 260
Exponentially decaying window, *see* Decaying window
Extrapolation, 474

Facebook, 184, 342
Fact table, 51
Failure, 21, 27, 40–42
Faloutsos, C., 401, 435
False negative, 86, 97, 225

False positive, 86, 97, 138, 225
Family of functions, 98
Fang, M., 236
Fayyad, U.M., 278
Feature, 264, 310–312
Feature selection, 444
Feature vector, 438, 478
Fetterly, D., 69
Fikes, A., 68
File, 21, 22, 207, 225
Filtering, 137
Fingerprint, 111
First-price auction, 291
Fixedpoint, 100, 190
Flajolet, P., 160
Flajolet-Martin Algorithm, 141, 393
Flow graph, 39
Fortunato, S., 400
Fotakis, D., 400
French, J.C., 278
Frequent bucket, 217, 219
Frequent itemset, 4, 200, 210, 212, 356, 437
Frequent pairs, 211
Frequent-items table, 212
Freund, Y., 482
Friends, 342
Friends relation, 50
Frieze, A.M., 127
Frobenius norm, 407, 421
Furnas, G.W., 434

Gaber, M.M., 18
Ganti, V., 127, 278
Garcia-Molina, H., 18, 198, 236, 278, 401
Garofalakis, M., 160
Gaussian elimination, 166
Gehrke, J., 160, 278
Generalization, 443
Generated subgraph, 355
Genre, 310, 322, 336
GFS, *see* Google file system
Ghemawat, S., 68, 69
Gibbons, P.B., 160, 401
Gionis, A., 127, 160
Girvan, M., 401
Girvan-Newman Algorithm, 349
Global minimum, 328
GN Algorithm, *see* Girvan-Newman Algorithm
Gobioff, H., 69
Golub, G.H., 434
Google, 162, 173, 288
Google file system, 22
Google+, 342
Gradient descent, 334, 371, 465
Granzow, M., 435
Graph, 43, 55, 341, 342, 378, 385
Greedy algorithm, 282, 283, 286, 290
GRGPF Algorithm, 264
Grouping, 24, 32, 35
Grouping attribute, 32
Groupon, 345
Gruber, R.E., 68
Guha, S., 278
Gunda, P.K., 69
Gyongi, Z., 198

Hadoop, 22, 69
Hadoop distributed file system, 22
Hamming distance, 65, 94, 102
Harris, M., 336
Harshman, R., 434
Hash function, 9, 77, 81, 86, 135, 138, 141
Hash key, 9, 298
Hash table, 9, 10, 12, 209, 216, 219, 220, 298, 300, 379
Haveliwala, T.H., 198
HDFS, *see* Hadoop distributed file system
Head, 390
Heavy hitter, 379
Henzinger, M., 127
Hierarchical clustering, 241, 243, 261, 324, 347
Hinge loss, 464
HITS, 190
Hive, 68, 69

Hopcroft, J.E., 391
Horn, H., 69
Howe, B., 68
Hsieh, W.C., 68
Hub, 190
Hyperlink-induced topic search, *see* HITS
Hyperplane, 459
Hyracks, 39

Identical documents, 116
Identity matrix, 405
IDF, *see* Inverse document frequency
Image, 131, 311, 312
IMDB, *see* Internet Movie Database
Imielinski, T., 236
Immediate subset, 228
Immorlica, N., 127
Important page, 162
Impression, 280
In-component, 167
Inaccessible page, 185
Independent rows or columns, 417
Index, 10, 379
Indyk, P., 127, 160
Initialize clusters, 253
Input, 55
Insertion, 93
Instance-based learning, 441
Interest, 204
Internet Movie Database, 310, 336
Interpolation, 474
Intersection, 31, 34, 38, 75
Into Thin Air, 309
Inverse document frequency, *see* TF.IDF, 8
Inverted index, 162, 280
Ioannidis, Y.E., 401
IP packet, 131
Isard, M., 69
Isolated component, 168
Item, 200, 202, 203, 306, 322, 323
Item profile, 310, 313
Itemset, 200, 208, 210

Jaccard distance, 90, 92, 98, 311, 477
Jaccard similarity, 72, 80, 90, 185
Jacobsen, H.-A., 68
Jagadish, H.V., 160
Jahrer, M., 339
Jeh, G., 401
Joachims, T., 482
Join, *see* Natural join, *see* Multiway join, *see* Star join, 381
Join task, 41

K-means, 252
K-partite graph, 345
Kahan, W., 434
Kalyanasundaram, B., 304
Kamm, D., 336
Kang, U., 401
Kannan, R., 434
Karlin, A., 284
Kaushik, R., 127
Kautz, W.H., 160
Kernel function, 471, 475
Key component, 135
Key-value pair, 23–25
Keyword, 289, 317
Kleinberg, J.M., 198
Knuth, D.E., 18
Koren, Y., 339
Kosmix, 22
Krioukov, A., 69
Kumar, R., 18, 69, 198, 401
Kumar, V., 18

Label, 342, 438
Lagrangean multipliers, 49
Landauer, T.K., 434
Lang, K.J., 401
Laplacian matrix, 362
LCS, *see* Longest common subsequence
Leaf, 350
Learning-rate parameter, 446
Leiser, N, 69
Length, 144, 385
Length indexing, 117
Leskovec, J., 400–402
Leung, S.-T., 69

Likelihood, 367
Lin, S., 127
Linden, G., 339
Linear equations, 166
Linear separability, 445, 449
Link, 31, 162, 176
Link matrix of the Web, 191
Link spam, 181, 185
Littlestone, N., 482
Livny, M., 278
Local minimum, 328
Locality, 342
Locality-sensitive family, 102
Locality-sensitive function, 97
Locality-sensitive hashing, 86, 97, 312, 477
Log likelihood, 372
Logarithm, 12
Long tail, 202, 306, 307
Longest common subsequence, 94
Lower bound, 59
LSH, *see* Locality-sensitive hashing

Machine learning, 2, 316, 437
Maggioni, M., 434
Maghoul, F., 18, 198
Mahalanobis distance, 259
Mahoney, M.W., 401, 434
Main memory, 207, 208, 216, 241
Malewicz, G, 69
Malik, J., 401
Manber, U., 127
Manhattan distance, 91
Manning, C.P., 18
Many-many matching, 111
Many-many relationship, 55, 200
Many-one matching, 111
Map task, 23, 25
Map worker, 26, 27
Mapping schema, 56
MapReduce, 15, 19, 22, 28, 175, 177, 227, 273, 381, 388, 456
Margin, 459
Market basket, 4, 16, 199, 200, 207
Markov process, 165, 168, 375
Martin, G.N., 160
Master controller, 23, 24, 26
Master node, 22
Matching, 285
Matias, Y., 160
Matrix, 29, *see* Transition matrix of the Web, *see* Stochastic matrix, *see* Substochastic matrix, 175, 190, *see* Utility matrix, 326, *see* Adjacency matrix, *see* Degree matrix, *see* Laplacian matrix, *see* Symmetric matrix
Matrix multiplication, 36, 37, 60
Matrix of distances, 415
Matthew effect, 14
Maximal itemset, 210
Maximal matching, 285
Maximum-likelihood estimation, 367
McAuley, J., 402
Mean, *see* Average
Mechanical Turk, 444
Median, 142
Mehta, A., 304
Melnik, S., 401
Merging clusters, 244, 247, 258, 262, 267, 271
Merton, P., 18
Miller, G.L., 401
Minhashing, 79, 89, 92, 99, 312
Minicluster, 256
Minsky, M., 482
Minutiae, 111
Mirrokni, V.S., 127
Mirror page, 73
Mitzenmacher, M., 127
ML, *see* Machine learning
MLE, *see* Maximum-likelihood estimation
Model, 367
Moments, 143
Monotonicity, 210
Montavon, G., 481
Moore-Penrose pseudoinverse, 427
Most-common elements, 155
Motwani, R., 127, 160, 236, 278

Mueller, K.-R., 481
Multiclass classification, 438, 453
Multidimensional index, 476
Multihash Algorithm, 220
Multiplication, 29, *see* Matrix multiplication, 175, 190
Multiset, *see* Bag
Multistage Algorithm, 218
Multiway join, 47, 381
Mumick, I.S., 160
Mutation, 96

Name node, *see* Master node
Natural join, 32, 35, 36, 46
Naughton, J.F., 69
Navathe, S.B., 236
Near-neighbor search, *see* Locality-sensitive hashing
Nearest neighbor, 442, 470, 479
Negative border, 228
Negative example, 445
Neighbor, 374
Neighborhood, 385, 393
Neighborhood profile, 385
Netflix challenge, 2, 308, 335
Network, *see* Social network
Neural net, 441
Newman, M.E.J., 401
Newspaper articles, 113, 299, 308
Non-Euclidean distance, 250, *see* Cosine distance, *see* Edit distance, *see* Hamming distance, *see* Jaccard distance
Non-Euclidean space, 264, 266
Norm, 91
Normal distribution, 255
Normalization, 319, 321, 332
Normalized cut, 361
NP-complete problem, 355
Numerical feature, 438, 478

O'Callaghan, L., 278
Off-line algorithm, 282
Olston, C., 69
Omiecinski, E., 236
On-line advertising, *see* Advertising
On-line algorithm, 16, 282
On-line learning, 443
On-line retailer, 202, 280, 306, 307
Open directory, 182, 444
OR-construction, 99
Orr, G.B., 481
Orthogonal vectors, 242, 408
Orthonormal matrix, 417, 422
Orthonormal vectors, 409, 412
Out-component, 167
Outlier, 241
Output, 55
Overfitting, 317, 334, 441, 442, 455, 479
Overlapping Communities, 367
Overture, 289
Own pages, 186

Paepcke, A., 127
Page, L., 161, 198
PageRank, 3, 16, 29, 30, 40, 161, 163, 175
Pairs, *see* Frequent pairs
Palmer, C.R., 401
Pan, J.-Y., 401
Papert, S., 482
Parent, 349
Park, J.S., 236
Partition, 359
Pass, 208, 211, 219, 224
Path, 385
Paulson, E., 69
PCA, *see* Principal-component analysis
PCY Algorithm, 216, 219, 220
Pedersen, J., 198
Perceptron, 437, 441, 445, 479
Perfect matching, 285
Permutation, 80, 85
PIG, 68
Pigeonhole principle, 355
Piotte, M., 339
Pivotal condensation, 405
Plagiarism, 73, 203

Pnuts, 68
Point, 239, 269
Point assignment, 241, 252, 348
Polyzotis, A., 68
Position indexing, 119, 120
Positive example, 445
Positive integer, 154
Powell, A.L., 278
Power Iteration, 405
Power iteration, 406
Power law, 13
Predicate, 316
Prefix indexing, 117, 119, 120
Pregel, 43
Principal eigenvector, 165, 405
Principal-component analysis, 403, 410
Priority queue, 247
Priors, 369
Privacy, 282
Probe string, 119
Profile, *see* Item profile, *see* User profile
Projection, 31, 34
Pruhs, K.R., 304
Pseudoinverse, *see* Moore-Penrose pseudoinverse
Puz, N., 68

Quadratic programming, 465
Query, 132, 151, 273
Query example, 471

R-tree, 278
Rack, 20
Radius, 249, 251, 385
Raghavan, P., 18, 198, 401
Rahm, E., 401
Rajagopalan, S., 18, 198, 401
Ramakrishnan, R., 68, 278
Ramsey, W., 303
Random hyperplanes, 103, 312
Random surfer, 162, 163, 168, 182, 374
Randomization, 224
Rank, 416
Rarest-first order, 299
Rastogi, R., 160, 278
Rating, 306, 309
Reachability, 387
Recommendation system, 16, 305
Recursion, 40
Recursive doubling, 389
Reduce task, 23, 25
Reduce worker, 26, 28
Reducer, 25
Reducer size, 52, 58
Reed, B., 69
Reflexive and transitive closure, 387
Regression, 438, 475, 479
Regularization parameter, 464
Reichsteiner, A., 435
Reina, C., 278
Relation, 31
Relational algebra, 30, 31
Replication, 22
Replication rate, 52, 59
Representation, 264
Representative point, 261
Representative sample, 135
Reservoir sampling, 160
Restart, 375
Retained set, 256
Revenue, 290
Ripple-carry adder, 154
RMSE, *see* Root-mean-square error
Robinson, E., 69
Rocha, L.M., 435
Root-mean-square error, 308, 327, 421
Rosa, M., 400
Rosenblatt, F., 482
Rounding data, 321
Row, *see* Tuple
Row-orthonormal matrix, 422
Rowsum, 264
Royalty, J., 69

S-curve, 87, 97
Saberi, A., 304
Salihoglu, S., 68
Sample, 224, 228, 231, 233, 253, 261, 265

Sampling, 134, 148
Savasere, A., 236
SCC, *see* Strongly connected component
Schapire, R.E., 482
Schema, 31
Schutze, H., 18
Score, 109
Search ad, 280
Search engine, 173, 189
Search query, 131, 162, 184, 280, 298
Second-price auction, 291
Secondary storage, *see* Disk
Selection, 31, 33
Sensor, 131
Sentiment analysis, 445
Set, 79, 116, *see* Itemset
Set difference, *see* Difference
Shankar, S., 69
Shawe-Taylor, J., 481
Shi, J., 401
Shim, K., 278
Shingle, 75, 89, 114
Shivakumar, N., 236
Shopping cart, 202
Shortest paths, 43
Siddharth, J., 127
Signature, 78, 81, 89
Signature matrix, 81, 86
Silberschatz, A., 160
Silberstein, A., 68
Similarity, 4, 15, 72, 199, 312, 320
Similarity join, 53, 59
Simrank, 374
Singleton, R.C., 160
Singular value, 417, 421, 422
Singular-value decomposition, 326, 403, 416, 426
Six degrees of separation, 387
Sketch, 104
Skew, 26
Sliding window, 132, 148, 155, 269
Smart transitive closure, 390
Smith, B., 339
SNAP, 400
Social Graph, 342
Social network, 341, 342, 403
SON Algorithm, 226
Source, 384
Space, 90, 91, 239
Spam, *see* Term spam, *see* Link spam, 344, 443
Spam farm, 185, 188
Spam mass, 188, 189
Sparse matrix, 29, 79, 81, 175, 176, 306
Spectral partitioning, 359
Spider trap, 168, 171, 191
Splitting clusters, 267
SQL, 20, 31, 68
Squares, 383
Srikant, R., 236
Srivastava, U., 68, 69
Standard deviation, 257, 259
Standing query, 132
Stanford Network Analysis Platform, *see* SNAP
Star join, 51
Stata, R., 18, 198
Statistical model, 1
Status, 299
Steinbach, M., 18
Stochastic gradient descent, 334, 469
Stochastic matrix, 165, 405
Stop clustering, 245, 249, 251
Stop words, 8, 77, 114, 203, 311
Stream, *see* Data stream
Strength of membership, 372
String, 116
Striping, 30, 175, 177
Strong edge, 344
Strongly connected component, 167, 391
Strongly connected graph, 165, 386
Substochastic matrix, 168
Suffix length, 121
Summarization, 3
Summation, 154
Sun, J., 435
Supercomputer, 19
Superimposed code, *see* Bloom filter, 159

Supermarket, 202, 224
Superstep, 44
Supervised learning, 437, 439
Support, 200, 225, 226, 228, 230
Support vector, 460
Support-vector machine, 437, 442, 459, 479
Supporting page, 186
Suri, S., 401
Surprise number, 144
SVD, *see* Singular-value decomposition
SVM, *see* Support-vector machine
Swami, A., 236
Symmetric matrix, 363, 404
Szegedy, M., 160

Tag, 312, 345
Tail, 390
Tail length, 141, 393
Tan, P.-N., 18
Target, 384
Target page, 186
Tarjan, R.E., 391
Task, 21
Taxation, 168, 171, 186, 191
Taylor expansion, 12
Taylor, M., 303
Telephone call, 344
Teleport set, 182, 183, 188, 375
Teleportation, 172
Tendril, 167
Term, 162
Term frequency, *see* TF.IDF, 8
Term spam, 162, 185
Test set, 442, 449
TF, *see* Term frequency
TF.IDF, 7, 8, 311, 441
Theobald, M., 127
Thrashing, 177, 216
Threshold, 87, 157, 200, 226, 230, 445, 451
TIA, *see* Total Information Awareness
Timestamp, 149, 270
Toivonen's Algorithm, 228
Toivonen, H., 236
Tomkins, A., 18, 69, 198, 401
Tong, H., 401
Topic-sensitive PageRank, 181, 188
Toscher, A., 339
Total Information Awareness, 5
Touching the Void, 309
Training example, 438
Training rate, 449
Training set, 437, 438, 444, 454
Transaction, *see* Basket
Transition matrix, 375
Transition matrix of the Web, 164, 175, 176, 178, 403
Transitive closure, 41, 387
Transitive reduction, 391
Transpose, 191
Transposition, 96
Tree, 246, 264, 265, *see* Decision tree
Triangle, 378
Triangle inequality, 91
Triangular matrix, 209, 218
Tripartite graph, 345
Triples method, 209, 218
TrustRank, 188
Trustworthy page, 188
Tsourakakis, C.E., 401
Tube, 168
Tuple, 31
Tuzhilin, A., 338
Twitter, 299, 342

Ullman, J.D., 18, 68, 69, 236, 278, 400
Undirected graph, *see* Graph
Union, 31, 34, 38, 75
Unit vector, 404, 409
Universal set, 116
Unsupervised learning, 437
User, 306, 322, 323
User profile, 314
Utility matrix, 306, 309, 326, 403
UV-decomposition, 326, 336, 403, 469

VA file, 476
Valduriez, P., 401
Van Loan, C.F., 434

Vapnik, V.N., 482
Variable, 144
Vassilvitskii, S., 401
Vazirani, U., 304
Vazirani, V., 304
Vector, 29, 91, 95, 165, 175, 190, 191, 240
Vigna, S., 400
Vitter, J., 160
Volume (of a set of nodes), 361
von Ahn, L., 313, 339
von Luxburg, U., 401
Voronoi diagram, 471

Wall, M.E., 435
Wall-clock time, 47
Wallach, D.A., 68
Wang, J., 336
Wang, W., 127
Weak edge, 344
Weaver, D., 68
Web structure, 167
Weight, 445
Weiner, J., 18, 198
Whizbang Labs, 2
Widom, J., 18, 69, 160, 278, 401
Wikipedia, 344, 444
Window, *see* Sliding window, *see* Decaying window
Windows, 12
Winnow Algorithm, 449
Word, 203, 240, 311
Word count, 24
Worker process, 26
Workflow, 39, 41, 45
Working store, 130

Xiao, C., 127
Xie, Y., 435

Yahoo, 289, 312
Yang, J., 401, 402
Yerneni, R., 68
York, J., 339
Yu, J.X., 127
Yu, P.S., 236
Yu, Y., 69

Zhang, H., 435
Zhang, T., 278
Zipf's law, 15, *see* Power law
Zoeter, O., 303