Index

A
additive logistic regression (ALR) 116, 125-126, 128
Agglomerative Hierarchical Clustering (AHC) 32, 36
application programming interface (API) 20
area under the curve (AUC) 123
Artificial intelligence (AI) 135

B
Boolean-valued information system 266-269

C
Car Evaluation (CE) 274, 276
Cascading Style Sheets (CSS) 141
centroids 156-157, 160-165, 167
document clustering 31-33, 35-39, 44-45, 48, 53
Document Type Description (DTD) 7

E
E-journals 226-228, 235, 247-248
Entity Search Term (EST) 8
textbook 188, 195

F
Finite State Machines (FSMs) 16
first normal form (1XNF) 66
frequently asked questions (FAQ) 139

G
Generalized Lloyd Algorithm (GLA) 156
Genetic Algorithm (GA) 158
Graphical Notation for Data Type Definition (GN-DTD) 55

H
Hadoop framework 263
Hidden Markov model (HMM) 81
hypertext 141, 186, 188-189, 193-194, 205-206, 239, 252, 258
Hyper Text Mark-up Language (HTML) 141
### Index

| **I** | IEEE International Conference on Data Mining (ICDM) 209  
|       | Information Retrieval System (IRS) 37  
|       | Information science (IS) 132, 137-138, 151  
|       | inheritance link 56, 64-65, 67-68  
|       | inverse document frequency (IDF) 214  
| **J** | Java Enterprise Edition (J2EE) 20  
| **K** | Keyword Query Graph (KQG) 3, 10  
|       | Keyword Query State Machine (KQSM) 3, 10  
|       | k-fold Cross Validation (k-fold CV) 216, 220  
|       | K-Means (KM) 157  
|       | K-Nearest Neighbor (KNN) 208  
|       | knowledge base (KB) 135  
| **L** | learning trees 22-23, 28  
|       | Legitimate Peripheral Participation (LPP) 133  
|       | lexicon-based approach 84  
|       | Life Insurance Promotion (LIP) 275  
|       | Lloyd’s K-Means algorithm 162-165  
|       | longest common prefix (LCP) 40  
|       | Lowest Common Ancestor (LCA) 3  
| **M** | Magazine Promotion (MP) 275  
|       | Management Information System Quarterly (MISQ) 210  
|       | MapReduce 250-255, 257-259, 261-265  
|       | Maximum Attribute Dependency (MDA) 277  
|       | Mereology 136-137  
|       | Mereotopology theory 137  
|       | Min-Min Roughness (MMR) 277  
|       | mobile originated (MO) 20  
|       | mobile terminated (MT) 20  
|       | Molecular Diversity Preservation International (MDPI) 237  
|       | Most Related Entity Instance Connecting Tree (MREIMCT) 9  
|       | Most Related Entity Instance (MREI) 8  
|       | multi-soft-sets 269  
|       | multi-tier web application 25  
| **N** | Naïve Bayes (NB) 208-209, 220  
|       | Natural Language Processing (NLP) 86  
|       | Nondeterministic Finite Automaton (NFA) 16  
|       | normalization algorithm 54, 63, 73-74, 76  
| **O** | object features 84  
|       | Object Oriented Database (OODB) 281  
|       | Ontocop 132-133, 138-144, 150-153, 155  
|       | Open Directory Project 260  
|       | Opinion Mining 78-80, 82, 86, 90, 94  
|       | opinion words 84, 86, 88-90, 93  
| **P** | PageRank 250-251, 253, 255, 264-265  
|       | parameterless k-means algorithm 158, 160  
|       | Part-of link 57-58, 64, 72  
|       | Pattern Reduction (PR) 159  
|       | Percentage Split (PS) 216, 220  
|       | Portable Document Format (PDF) 239  
|       | pre-processing 33, 38, 86, 90, 160, 163, 274  
|       | Pruning Phase 121  
|       | Pubic Library of Science (PloS) 237  
| **Q** | query log 46-52  
|       | query recommendation 46-49, 52-53  
| **R** | reading behaviours 188-190  
|       | reading process 188-193, 200, 202, 204-205  
|       | Relational Database (RDB) 281  
|       | reverse arithmetic encoding 98, 114  
|       | Root-based Stemming 211  
|       | Rule Based classifier 116, 118  
|       | Rule Growing 121  
| **S** | search engines 31-33, 35-38, 46, 48, 52-53, 79, 151, 169-170, 207, 245-246, 250, 258, 264  
|       | second normal form (2NF) 62  
|       | Sentence-level sentiment classification 82  
|       | sentiment analysis 78, 80-83, 93-94  
|       | Sequential Minimal Optimization (SMO) 209, 215, 220
Index

short message service (SMS) 19-20
Simple API for XML (SAX) 109
Simplified DTD Entity-Attribute Graph (SDEAG) 7
Smallest Lowest Common Ancestor (SLCA) 3
Social Bookmarking Site (SBS) 170
Social Semantic Ranking (SSR) 170, 172, 185
Social Semantic Relevance (S2R) 170, 175, 185
stack based sort-merge algorithm (DIL) 4, 16
stemming 34, 36, 38, 85, 207-209, 211-214, 216, 218-224
stop words 34, 38, 86-87, 213-214, 221
structural identifiers 99
Structured Query Language (SQL) 282
Structure Index Tree (SIT) 99
structure summary tree 99
Student Enrolment Qualifications (SEQ) 274
subjectivity classification 82
Suffix Tree Clustering (STC) 33, 36
Suffix Tree Document (STD) 36
Support Attribute Representative (SAR) 266-267, 270-271, 279
Support Vector Machine (SVM) 207-209

T

tacit knowledge 132, 134-135, 141
Tanimoto coefficient 49-51
taxonomy 132, 136-138, 146, 150, 172
Term Frequency Inverse Document Frequency (TFIDF) 214
Term Frequency (TF) 214

Text classification 81, 94, 207-212, 221-225
tokenization 34, 222
Total Roughness (TR) 277
transitively dependent 61-62, 66

V

Vector Quantization (VQ) 156
vector space document (VSD) 36
Vector Space Model 31-32, 34, 44, 170, 175, 186
virtual community 132, 134, 141-142, 153-155

W

Watch Promotion (WP) 275
Web crawlers 258
Web Personalization 260

X

XGrind 98, 115
XLCA Connecting Tree (XLCACT) 1, 4
XMll 97-98, 114
XMLCQ-Compressor (XMLCQ-C) 100
XMLCQ-Decompressor (XMLCQ-D) 100
XMLCQ-Query Processor (XMLCQ-QP) 100
XML documents schema (DTD) 75
XML documents (XNF) 72
XML functional dependency (XFD) 72
XML stream system (XSS) 1