Index

A
advanced data cleansing 218
affinity 148
affinity measure 158
aggregate usage profiles 283
AISEC algorithm 156
antigen presenting cell 151
artificial immune system 145
asset mentality 7
association rule mining 34
association rules 359
attribute grammars (AGs) 169, 172
attribute lists 191
auto-immune response 150
automatic graph layout 256

B
backward-link count 53
backward reference 324
Bayes theorem 153
Bayesian classification 152
Bee-gent system 239
bootstrapping 191

C
canonical hierarchy 57
cell death process 157
chi-square test 197
CiteSeer 282
classification 9, 34
click-stream 33
clonal selection 149
clustering 9, 34, 113, 357
collaborative filtering (CF) 277
content mining 394
context attributes 178
context graph 52, 58
converging rate 101
cookie 341
costimulation 157
customer profile exchange (CPEX) 39
customer relationship management (CRM) 34, 374
cyberspace 253
cyberspace navigation 253

D
Danger Theory 145
data analysis 345
data cleansing 208, 214, 375
data collection 345
data cube 189
data definitions 15
decision-making 6
decision trees 200
demographics 30
disambiguate polysemous terms 83
distributed software architecture 355
document type definitions (DTDs) 170
domain knowledge representation 286
domain ontology 71, 280
dynamic media language 130
E
e-learning 238
electronic commerce 373
e-mail classification 154
empirical performance 107
evaluated attributes 178
expansion algorithm 101
extensible markup language (XML) 39, 169
F
fit tests 219
focused crawler 53
forecasting 9
forward-link count 53
forward references 326
fuzzy deixis 130
G
gap sessions 329
gene libraries 157
generated code 170
graph clustering 258
H
HITS algorithm 407
homogeneity 401
humos/wifs 36
hyperlinks 208
hypertext induced topic search 55
hypertext markup language (HTML) 122
I
IC systems 232
independent TLD target distribution (ITTD) 218
IndexFinder 38
information customization 231
information foraging theory 28
information retrieval 211
informavores 28
interlinking pattern 221
interval sessions 328
inverse documents frequency 57
inverted files 309
J
Java server pages (JSP) 126
Java technology 355
K
knowledge base construction 281
knowledge discovery 393
knowledge-enhanced Web data mining 281
knowledge management 1, 9
L
language 111
Letizia 36
logname 374
lymphocytes 148
M
MA algorithm 113
MASEL algorithm 316
memory 111
memory cells 150
metadata 3
metadata management 1
METALA architecture 362
mining traversal patterns 376
modeling users 232
monosemous terms 82
multi-agent system 235
multi-agents 228
multiple site link structure analysis 208
multiplicative adaptive refinement search (MARS) 99

N
naïve Bayesian classifiers 152, 199
negative selection 150
New Testament semantics 183
New Testament SLXS specification 183
newsgroup 393
newsgroup mining structure 405
non-text media 119

O
online analytical processing (OLAP) 189
OntoLearn system 73
ontology 279
ontology-based personalization 294
ontology engineering 280
ontology learning 69
open profiling standard (OPS) 39

P
page crawlers 126
page specification languages 126
page view 32, 33
PageRank 54
path analysis 9
path completion 375
pattern analysis 34
pattern detection 197
pattern discovery 34, 290, 404
personal WebWatcher 37
personalization 28, 30, 42, 113, 276, 393
personalized information description language 42
personalized output 35
platform for privacy preferences (P3P) 39
precision performance 109

Q
query 341
query answer time 315
query distribution 312
QueryConstructor 106

R
RankPower 110
recall performance 109
recognition region 148
referring phrases 128
reinforcement learning 58
relevancy-context-graph model 60
remotehost 374
resource description framework (RDF) 39
results pages 344
robust fuzzy C-medoids algorithm 369
root set 55

S
search engine 229, 339
search engine usage 120, 212, 309, 339
selective superiority problem 366
semantic awareness 69
semantic knowledge 276
semantic pattern 83
semantic similarities measurement 291
semantic web 69
semantic web mining 280
sequential patterns 34, 361
server session 32
session identification 33, 375
sessionization 379
“shallow” generalized suffix trees 331
site mapping 253
specification language for XML semantics (SLXS) 170
statistics tree (ST) 193
structure mining 394
system engineering benefits 70

T
term frequency 57
text categorization 52
textual similarity 56
time stamp 341
topic-specific crawling system 50
topic-specific spider 53
topological cleansing 217
topological community mining 222
training data set 189
transaction identification 376
transactionization 380
traversal path patterns 322

U
Ukkonen algorithm 322
upper domain ontology (UDO) 70
usage mining 394
usage patterns 37
Usenet site 396
user identification 33, 375
user queries 316
username 374

V
validation 220
viewing patterns 339

W
web-based warehouse 15
web “crawler” 126, 211
web documents 344
web graph clustering 253
web graph displays 254
web growth 210
web logs 309, 322, 398
web mining 119, 145, 237, 278, 322, 339
web page captions 121
web personalization 27, 277
web robot 52
website topology 367
WebTrie 382
web usage mining 276, 307, 355, 373, 375
web user 322
web utilization miner 38
web warehousing 1
WebWatcher 36

X
XML semantics 169

Z
Zipf’s law 308