Index

A
Abstraction, 12–13, 18, 243, 244
Actionable information, 6–7, 9, 10, 24
Adaboost method, 284–285, 293, 296
Adjusted Stability Measure (ASM), 116
Advanced message queuing protocol (AMQP), 250
Aggregation, 11–12
Amazon’s elastic block storage, 245
Analysis block controller (ABC), 241, 248, 249, 251, 253
Analysis of variance (ANOVA) analysis, 127–128
Annotated pairing
  algorithm, 38–39
  processing pipelines, 32–33
  rules, 37–38
Augmented summary graph, 165

B
Bi-level indexing keyword search approach (BLINKS)
  backward expansion, 165
  blocks, 164–165
  complexity, 167–169
  computing time, 169
  intra-level block index, 165
  iterations, 168
  node-keyword map, 165
  partitioning, 178, 179
  portals, 165
  query response, 169
  scoring function, 165
  search algorithm, 167
  vs. SearchWebDB, 173
  single-level index, 165
solutions, 179, 180
top-level block index, 165
visitNode procedure, 167
vs. YAANII, 174
Binary-class classification
classification, 283
clustering-based framework, 294 (see also Clustering-based subspace modeling method (CLU-SUMO))
negative data instances, 282
positive data instances, 282
BLINKS. See Bi-level indexing keyword search approach (BLINKS)
Business process optimization. See Process and operational data combination

C
Canonical correlation analysis (CCA)
collinearity, 225
dataset size, 223–225
definition, 222
disadvantages, 222
linear correlations determination, 222
objective, 222
ranking of correlated features, 225
regularized-type CCA, 222–223
ridge-type CCA, 222
Categorization, 12–13
Certain data, 51–52
Cloud architecture, 243
Clustering-based subspace modeling method (CLU-SUMO)
classification phase, 290
enhancements, 286
experimental results, 294–296
experimental setup, 292–294
Index

function definitions
- PCP, 289
- score, 290
- SVD, 289
- Z function, 289

integrated subspace modeling and classification, 288
k-means clustering, 285, 288
learning phase, 290
negative group centroid, 291
non-target concept class, 287
procedures, 286–287
ranking scores, 292
semantics information and clustering, 285
testing data instances, 292
training data instances, 291
utilization class label information, 288

Collaborative forensic framework (CUFF)
advantages, 238
analysis block
- acquisition, 248
- AMQP, 250
- analysis node agents, 251–252
- anti-starvation requirement, 252
- controller, 241, 248, 249
- data encapsulation method, 249
- discrimination, 249
- JSON, 250
- processing resources, 241, 248
- RabbitMQ, 250, 251
- scheduling utilities, 249, 250
- scheduling utilities evaluation, 253–255
- validation, 248
- cloud architecture, 243
- COTS software, 242
- Cuff Link, 242
- DNS server, 244
- functionality, 243
- inter-component communication, 244
- key functions, 243
- REST web services, 244
definition, 238
desktop applications, 243
infrastructure, 239
storage
- Amazon’s elastic block storage, 245
- DFXML file, 247
- disk images tracking, 241
- OpenStack architecture, 245
- persistent data storage, 245
- response time, 242
- Swift, 245
- UUID, 245–246
- VM images, 245

user interface, 242, 255–257
virtual machine, 243
Commercial off-the-shelf (COTS) software, 242
Confidence-based top-k join query (PTopkJQ), 56
Constructive cost model (COCOMO), 207
Constructive cost model II (COCOMO-II), 214–216
Construct validity, 218
Content management module, 206
Cosine similarity (CSim), 86
CostDTTree, 296
Cross-Origin Resource Sharing (CORS), 150
CUF. See Collaborative forensic framework (CUFF)
C3W method, 137

D
Databases, definition, 5
DataBoost-IM method, 285
Data integration
- heterogeneous semantics, 15
- heterogeneous sources, 14–15
- human capabilities, 18
- incommensurate information, 17
- knowledge maintenance, 21
- knowledge modeling, 19–20
- knowledge sharability, 20
- mediating knowledge, 17
- mismatches, 15–16
- payroll file, 16
- personnel file, 16
- public data sources, 16
- redundancy removal, 13
- trusting, 20–21
- uncertainty, 18–19
Data warehousing, 10
DBXplorer, 198
2D discrete Fourier transform (2D-DFT), 76
Dense test, 66–67
DesignCPN, 97
1D-fast Fourier transform (1D-FFT), 76
Different test, 66–67
Digital forensic examinations, 237–238
DISCOVER, 198
Discrete Fourier transform (DFT), 74
Discrete wavelet transform (DWT), 74
Distributed processing, forensics, 240
2D mapping indexes, 55–56
2D Mellin-Fourier transform (2D-MFT), 74
2D modified Mellin-Fourier transform (2D-MMFT)
advantages, 93
IPD (see Inverse pyramid decomposition (IPD))
RSTC description, 75
Document object model (DOM) tree, 197
Domain Name System (DNS) server, 244

E
Exponential polar transform (EPT), 78
External interval tree index, 53–54

F
Feature selection techniques
defect prediction models, 114
feature ranking, 117 (see also Filter-based feature ranking techniques)
feature subset selection, 115
filters, wrappers, and embedded methods, 115
stability
advantage, 114
ANOVA analysis, 127–128
average stability, 124–126
dataset perturbation, 122–123
datasets, 121–123
definition, 116
degree of perturbation impact, 127
measurement, 123–124
threats to validity, 128–129
Filter-based feature ranking techniques
categories, 116–117
commonly used feature ranking techniques, 117–118
feature score, 116
signal-to-noise (S2N), 121
threshold-based feature ranking techniques
algorithm, 118, 119
area under precision-recall curve (PRC), 121
area under ROC (AUC), 120
deviance (Dev), 120
false positive and negative rates, 118
F-measure (FM), 118–119
geometric mean (GM), 120
Gini index (GI), 120
Kolmogorov-Smirnov (KS), 120
mutual information (MI), 120
odds ratio (OR), 119
posterior probabilities, 118
power (PO), 119
probability ratio (PR), 119
ture positive and negative rates, 118
Forensics standardizations, 240
Formalisms integration, 95, 96
Formal verification, 95

G
Garfinkel’s digital forensics XML (DFXML) file, 247
Geographic aggregations, 12

H
High-level mediators, 4
Histogram descriptor, 74
Homonyms, 15
HTML-aware tools, 197

I
Imbalanced multimedia data
binary-class classification (see Binary-class classification)
boosting methods, 284–285
CLU-SUMO (see Clustering-based subspace modeling method (CLU-SUMO))
cost sensitive learning methods, 285
data sampling, 283–284
learning, 283
searching, indexing and retrieving, 282
Information, definition, 5–6
Information integration, 29
Intel MashMaker, 137
Internet Scrapbook, 137
Invariant object representation
closest objects search, image databases, 83–84
closest vector search, image database CSim, 86
modified squared Euclidean distance, 85–86
MSCSim, 87
R-dimensional vectors distance, 84
scalar product, 84
SCSim, 87
2D-MMFT
advantages, 93
IPD (see Inverse pyramid decomposition (IPD))
RSTC description, 75
2D rigid transforms, 74
experimental results
closest images, 88–90
content-based object retrieval, 87, 88
discretization grid of LPT, 87, 88
distance representation, 90–92
3D object detection, 90
multi-view images, 90, 91
points retained by EPT and LPT, 87, 88
scanned texts documents, 92–93
test image after EPT, 87, 89
objects descriptions, 74
RST-invariant descriptors, 74–75
Inverse pyramid decomposition (IPD)
amplitude spectrum, 79, 80
bi-polar transformation, 75
complex coefficients, 79, 80
2D-DFT coefficients, 76
denormalized coefficients, 80
difference image, 82
EPT, 78
Fourier coefficients, 76–77, 79
intermediate spectrum coefficients, 76
interpolated coefficients, 81
inverse 2D-DFT, 81, 82
inverse EPT, 81
IPD-MMFT block diagram, 82, 83
LPT, 77–78
modules and phases, 77
R-dimensional RSTC-invariant vector, 79, 81
second 2D-DFT, 78–79
IPD. See Inverse pyramid decomposition (IPD)

J
JavaScript object notation (JSON), 250
Join query, 56
JOUS-Boost, 285

K
Kendall’s ranking correlation, 228
Kendall’s tau scores, 228
Keywords search based systems
BLINKS (see Bi-level indexing keyword search approach (BLINKS))
comparison with relations, 172–173
data structure matching, parts retrieval, 163
effectiveness evaluation, 184–185
efficiency evaluation, 184
notation, 167
ontology of reference, 177, 178
rank candidate solutions, 164
ranking, 174–177
SearchWebDB (see SearchWebDB)
YAANII (see YAANII)
KLOC software metric, 207
Knowledge, definition, 6

L
Left boundary leaf, 61
Linear Temporal Logic (LTL), 108–109
Local principal components and Kendall’s ranking (LPC-KR)
between-groups correlations, 225
compute Kendall based between groups correlations, 228
compute within group correlation, 227
fit linear model, 228
Kendall’s ranking correlation, 228
Kendall’s tau scores, 228
linear models construction, 223, 226
multidrug resistance data, 230–233
multi-’omics’ analysis, 233
nutrigenomic dataset, 229–231
omics datasets, 223
threshold $\beta_W$, 227
within-group correlated features, 223
Logistic regression model, 294
Log-polar transform (LPT), 74, 77–78
Low-level mediators, 4
LPC-KR. See Local principal components and Kendall’s ranking (LPC-KR)

M
Manual filtering, 43–44
Mediators
architecture, 2–3
artificial intelligence technology, 24
conceptual principles, 7–8
automation, 9, 24
incompatibility delegation, 8
integrated actionable information, 9
one-directional flow, 8
semantics limitation, 8–9
vs. data warehousing, 10
definitions
actionable information, 6–7
databases, 5
information, 5–6
knowledge, 6
efficiency and reliability, 23
functionalities, 11
high-level mediators, 4
integration (see Data integration)
interfaces, 4
low-level mediators, 4
mediation technology state, 24
motivation, 3–4
operations and volume management
abstraction, 12–13
aggregation, 11–12
information overload, 10
ranking, 13–14
redundancy removal, 13
selection and projection, 11
volume reduction, 10
partitioning vs. centralization, 22
private vs. public mediation, 22
role of mediation, 2
search ad discovery, 5
security and privacy, 22–23
Memory-loaded threshold interval index (MTII)
branching parameter, 64
leaf interval, 64
primary tree, storage space, 63–64
secondary structures, 64, 65
stab 'n grab search, 65
uncertain objects, 64
x-bound tree, 64–65
Microblogging services, 264–266
Minimum bounding rectangle (MBR), 54–55
Minimum squared Euclidean distance, 83, 84
Mixup, 137
Modified squared cosine similarity (MSCSim), 87
Modified squared Euclidean distance (MEUD), 85–86
MTII. See Memory-loaded threshold interval index (MTII)
Natural Language Parsing (NLP), 188–190
Normalized discounted cumulative gain (NDCG), 195–196
Object Constraint Language (OCL), 96–98, 102, 103, 107–109, 157
Object Petri net (OPN)
advantage, 96
association end specification, 102–104
initialization approach
LTL properties, 108–109
object and sequence diagrams, 104–105
object distribution, 105–107
on-the-fly tester approach, 107
PROD, 107, 108
reachability graph inspection approach, 107
LEMMA, 97
mapping algorithm, 98
marking approach, 96
OCL invariants, 96–97
peer, 101, 102
peer to peer class diagram, 99
state machine
dynamic model, 101
n-tuples, 100
and sequence diagrams, 97, 99
simple and super transition models, 100
static and dynamic marking, 101
textual and graphical formalization, 97
Ontology-based tools, 197
OpenStack architecture, 243, 245
OPN. See Object Petri net (OPN)
Oversampling, 284
Partial least squares regression (PLS), 222
Partially-annotated pairing
algorithm, 40–41
processing pipelines, 32–33
rules, 40
Principal component projection (PCP)
function, 289
Private mediation, 22
Probabilistically constrained regions (PCRs), 57
Probabilistic distribution R-tree (PDR-tree), 57
Probabilistic inverted index, 57
Probabilistic join query (PJQ), 56
Probabilistic threshold join query (PTJQ), 56
Probability distribution functions (PDFs), 52
Probability threshold index (PTI), 54–55
Process and operational data combination, 30
advantages, 44–45
annotated pairing
algorithm, 38–39
processing pipelines, 32–33
rules, 37–38
business performance management, 31
experimental results, 45–47
experimental setup, 45
information integration, 29
input and output models, 34–35
manual filtering, 43–44
partially-annotated pairing
algorithm, 40–41
processing pipelines, 32–33
rules, 40
process mining, 31
process model content and structure, 30
sample scenario
  ContractNegotiation, 35
  CustomerData, 35
input and output sets, 36–37
inputData, TaskVar and ServiceInfo variables, 36
RentalCarSelection scenario, 35, 36
tree components, 36
schema/web service matching, 31–32
structure-based filtering algorithm, 42–43
processing pipelines, 33
rules, 41–42
Public mediation, 22

R
RabbitMQ, 250, 251
Range query process evaluation
grabs, 62
left stab, 61
right stab, 61–62
time bounds, 62–63
externalization, 63
TII structure
primary tree, 58–59
secondary structures, 59–60
x-bounds, 60–61
Ratio test, 68–70
READFAST (RF) architecture
architecture, 190–191
automatic extraction, hierarchical index, 189
browser and navigation index, 188
browsing interface, 189
evaluation purposes, 192
hierarchical document summary, 189
index extraction, 191
NLP, 188–190
search relevance evaluation, 194–196
standard measures, 192
system entity, 189
term extraction, 193–194
Redundant array of independent disks (RAID) storage systems, 240
Region descriptor, 74
Representational State Transfer (REST) web services, 244
ResampleLG method, 294, 296
Re-sampling, 294
Rich Internet application (RIA) technologies, 134
Ridge-type canonical correlation analysis (RCCA), 222
Right boundary leaf, 61
Rotation-scale-translation (RST), 74
S
Same-origin policy (SOP), 141
Same test, 66–67
Schema annotation, 33
Score Combination module, 290
SearchWebDB
augmented summary graph, 165, 180, 181
vs. BLINKS, 173
complexity, 169, 170
conjunctive queries, 165
element-to-query mapping, 165
scoring functions, 165–166
TOP-N function, 169
vs. YAANII, 173–174
Server-side logic processing, 134, 135
Shape boundary descriptor, 74
Signal-to-noise (S2N), 121
Similarity join query, 56
Simple multi-attribute rating technique (SMART)
alternatives courses of action identification, 204–205
COCOMO, 207–208
COCOMO-II, 214–216
construct validity, 218
current web information system, 204
data collection and analysis, 212
decision maker identification, 204
e external validity, 218
internal validity, 218
multi criteria decision analysis, 217
normalized weights of benefits, 212
 objective, 202
provisional decision and sensitivity analysis, 213–214
relevant attributes identification, 205–206
reuse alternative, weights of benefits, 208, 209
SRo, 203–204
stages, 203
swing weights technique, 209, 212
value and utility, 203
Singular value decomposition (SVD) function, 289
Skyline, 57
SMART. See Simple multi-attribute rating technique (SMART)
SMOTEBest, 285
Sparse test, 66–67
Spearman rank correlation coefficient, 116
Squared cosine similarity (SCSim), 87
Stabbing queries, 54
Stab 'n grab search, 61, 62, 65
State machine
dynamic model, 101
n-tuples, 100
of peer, 100
and sequence diagrams, 97, 99
simple and super transition models, 100
static and dynamic marking, 101
Structure-based filtering
algorithm, 42–43
processing pipelines, 33
rules, 41–42
Swing weights technique, 209
Synonyms, 15

T
Temporal aggregations, 12
Threshold-based feature ranking techniques (TBFS)
algorith, 118, 119
area under precision-recall curve (PRC), 121
area under ROC (AUC), 120
deviance (Dev), 120
false positive and negative rates, 118
F-measure (FM), 118–119
geometric mean (GM), 120
Gini index (GI), 120
Kolmogorov-Smirnov (KS), 120
mutual information (MI), 120
odds ratio (OR), 119
posterior probabilities, 118
power (PO), 119
probability ratio (PR), 119
ture positive and negative rates, 118
Threshold interval indexing, 52
Threshold test, 67–68
Time Petri Nets, 98
Travel mashup application, 155
TRIO temporal logic, 97
experimenal results
object spread, 66–67
probability threshold, 67–68
query interval size, 68–70
test model, 65–66
external interval tree index, 53–54
high dimensionality, 57–58
indexing categorical data, 57
join query, 56
moving objects, 58
MTI, 63–65
PDFs, 52
problem statement, 53
processing range queries (see Range query process)
PTI, 54–55
ranking queries, 56
skyline searches, 57
threshold interval indexing, 52
x-bounds, 52
Uncertain discrete attribute (UDA), 57
Undersampling, 284
Universally unique identifier (UUID), 245–246

V
Value tree, 205
Variance-based clustering, 54
Video sharing
counter-terrorism, 272
education, 268–269
financial news, 271
gotracking, 269–270
marketing/ads, 269
medicine, 270–271
political, 272–273
self-expression/art, 266–268
Virtual machine, 243

W
Web Application Description Language (WADL), 152–153
Web functionality search engine, 141–143
Web Service Definition Language (WSDL), 152–153
Web 2.0 services, 261–263
Web Services for Remote Portlets (WSRP)
developed parts, sharing and reusing, 139
flexible integration, 140
integrated software enhancement, 139–140
interoperating components, 140–141
SOP, 141
web functionality/component retrieval method, 137–138
Web-user interface component
client-side scripting, 134, 135
C3W method, 137
integration and reuse
access tag, 150–151
copyright, 158–159
emulation, 145
extraction, 145
format of entity, 143–144
functionality interaction, 148
iframe tag, 150
integration testing, 156–157
interoperation model, 148–149
JavaScripts separation, 149–150
layout personalization, 154
maintainability of integrated application, 146–147
page transition and entity definition, 151–152
reliable and secure composition, 154–156
retrieval system, 141–143
script separation, 148
service granularity, 157–158
start page and optionlist components, 152, 153
UI component isolation, 148
usage control, 148
WADL, 152–153
WSDL, 152–153
Intel MashMaker, 137
Internet Scrapbook, 137
Mixup, 137
option list, 134, 135
server-side logic processing, 134, 135
text input field, 134, 136
web mashup technologies, 137
WSRP
developed parts, sharing and reusing, 139
flexible integration, 140
integrated software enhancement, 139–140
interoperating components, 140–141
SOP, 141
web functionality/component retrieval method, 137–138
Weight-balanced B-tree, 53
World Wide Web (WWW)
blogs, 263–265
microblogging services, 264–266
social networks
education, 274–275
health, 276
music, 275
researchers visualization, summary table, 277
sports, 273–274
statistical data, 262
video sharing
counter-terrorism, 272
education, 268–269
financial news, 271
geotracking, 269–270
marketing/ads, 269
medicine, 270–271
political, 272–273
self-expression/art, 266–268
Web 2.0 services, 261–263
Wrapper induction tools, 197
WSRP. See Web Services for Remote Portlets (WSRP)

Y
YAAIIN
BLINKS, 174
characteristics, 166
clustering, informative paths, 166, 171, 181
complexity, 170–172
pattern formula, 166
query solutions, 181–183
vs. SearchWebDB, 173–174