

Index

A

adaptive systems, 16–17
 advanced analytics. *See also*
 analytics
 AI and, 89
 cognitive computing, 88
 customers and, 90
 data mining and, 92–93
 descriptive, 87–88
 image analytics, 101–103
 machine learning, 88, 93–98
 maturity levels, 88
 open source tools, 105–106
 predictive, 88
 business value, 98–99
 use cases, 99
 prescriptive, 88
 SLA and, 91
 speech analytics, 103
 statistics and, 92
 text analytics, 99–101
 value and, 103–105
 AEMS (Area Energy
 Management Solutions),
 209–210
 AI (Artificial Intelligence)
 American Express, 10
 analytics and, 89
 games, 7
 Logic Theorist, 8
 McCarthy, John, 8
 Minsky, Marvin, 8
 Newell, Allen, 8
 NLP and, 10

Samuel, Arthur Lee, 7–8
 Simon, Herbert A., 8
 Turing, Alan, 6–7
 Weiner, Norbert, 7
 alternative models, 249
 ambiguity, resolving, 47–48
 American Express project, 10
 analytics, 4
 anticipatory in application
 building, 164–166
 maturity stages, 132–133
 APIs (Application
 Programming Interfaces),
 134
 application building
 anticipatory analytics,
 164–166
 corpora, 168–171
 domain, 160–161
 emerging platform, 158–159
 governments, 197–198
 city characteristics, 199–204
 community infrastructure,
 217–218
 data ownership, 205–206
 data value, 205–206
 healthcare, 212–213
 Internet of Everything,
 204–205
 open data movement, 204
 system of systems, 199
 technology, 206–212
 transportation
 infrastructure, 213–215

workforce and, 215–217
 healthcare
 approaches, 181–182
 Big Data analytics, 180–181
 clinical teaching and,
 193–195
 corpus, 184, 185
 data service providers, 179
 EMR (electronic medical
 record), 191–193
 foundations, 176–177
 government agencies, 179
 health improvement, 186–191
 ontologies, 182–183
 patients, 178
 patterns, 179–180
 payers, 179
 pharmaceutical companies,
 178
 providers, 178
 questions, 183–185
 system training, 185
 insights, data sources,
 166–168
 objectives, 159–160
 questions, 162–164
 testing, 171–173
 training, 171–173
 users, 161–162
 applications, cognitive
 systems, 20
*Artificial Intelligence at MIT:
 Expanding Frontiers*
 (Minsky), 78

ASR (automated speech recognition), 103
 ATM (Automated Teller Machine), 10
 audio, analyzing, 16
 Automatic Thinking, 13–14

B

BI (business intelligence),
 cloud computing and, 119
 biases, 13–14
 Big Data, 4
 architecture, 57
 databases, 58–61
 physical foundation, 58
 security infrastructure, 58
 compressed sensing, 63
 data services, 61
 data warehouses, 61–62
 databases, 59
 document databases, 60
 graph, 60
 KVP (Key-Value Pair), 59
 Polyglot Persistence, 60
 PostGIS/OpenGEO, 60
 spatial, 60
 description, 56
 dimensionality reduction, 62
 Hadoop, 63–67
 integration with traditional, 69
 numerical linear algebra, 62
 sketching and streaming, 62
 tool, 61
 variety, 56–57
 velocity, 57
 veracity, 57
 volume, 56
 business
 APIs (Application Programming Interfaces), 134
 change and, 125–126
 disruptive models, 126–127
 knowledge and, 127–128
 planning, 131–133
 markets, 136
 business-specific solutions, 134–135

C

call centers, 231
 categorization, 15
 chess notation, 76–77

“Chess Playing Programs and the Problem of Complexity” (Newell, Shaw, Simon), 8
 choice, 12–13
 city planning, 200–201
 classification, 75
 supervised learning and, 30–31
 cloud computing, 110
 BI (business intelligence), 119
 clinical research, 116–117
 data integration, 122
 data management, 122
 delivery models, 117–120
 distributed processing, 111–112
 elasticity, 111
 governance, 121–122
 hybrid cloud, 115–116
 IaaS (Infrastructure as a Service), 117
 containers, 118
 SDE, 118
 virtualization, 117–118
 models, 112–117
 MSPs (Managed Service Providers), 114–115
 multitenant cloud, 112–113
 necessity, 110
 PaaS (Platform as a Service), 120
 private cloud, 114
 public cloud, 112–114
 SaaS (Software as a Service), 118–120
 scaling, 111
 security, 121–122
 self-service provisioning, 111–112
 workload management, 120–121
 clustering, 97
 cognition, 11–12
 choice and, 12–13
 computer science and, 13
 judgment and, 12–13
 System 1 (Automatic Thinking), 13–14
 System 2 (Controlled, Rule-centric, Concentrated Effort), 14–15
 cognitive computing
 analytics, maturity and, 88
 domains of use, 5–6

as new generation, 2
 principles, 1–2
 cognitive science, 11–12
 cognitive systems
 adaptive, 16–17
 applications, 20
 business, 131
 APIs, 134
 planning, 131–133
 business and, 128–129
 components, 22–23
 continuous machine learning, 19
 corpus, 18, 23–26
 data
 analytics services, 28
 sources, 26–27
 data access, 18
 extraction services, 27–28
 data analytics services, 18–19
 data catalogs, 18
 deployment, 17–18
 development, 72–73
 features, 3–4
 hypotheses, 33–34
 generation, 34–35
 scoring, 35–36
 infrastructure, 17–18, 37
 learning process, 19
 machine learning and, 29–33
 data patterns, 29
 management services, 18
 metadata, 18
 NLP
 business problems, 50–53
 construction grammars, 46
 context, 40–42
 disambiguation, 47–48
 discourse analysis, 46–47
 history, 43
 language identification, 43
 lexical analysis, 45
 linguistics, 43
 Markov Models, 48–49
 morphology, 44
 phonology, 44
 pragmatics, 47
 syntax, 45–46
 tokenization, 43
 words, connecting, 42–43
 WSD, 49
 persistence, 84–85
 presentation services, 20, 36–37
 relationships, 15–16

- state, 74–75, 84–85
- taxonomies, 18, 25
- uses, 2–3
- visualization services, 20, 36–37
- columnar databases, 60
- compressed sensing, 63
- computer science, 11–12
- cognition and, 13
- “Computing Machinery and Intelligence” (Turing), 6
- concentrated effort in thinking, 14–15
- confidence estimation, Watson (IBM), 153–154
- construction grammars, 46
- continuous machine learning, 19
- Controlled, Rule-centric, Concentrated Effort, 14–15
- controlled thinking, 14–15
- coordination services, 61
- COPLink project, 208
- corpora, 18, 21
- in application building, 168–171
- regulatory considerations, 25–26
- security, 25–26
- Watson (IBM), 145–147
- customers
 - analytics and, 90
 - machine learning and, 90
 - NLP and, 51–53
- cybernetics, 7
- Cybernetics or Control and Communication in the Animal and the Machine* (Weiner), 7
- D**
- dark data, 68–69
 - application building, 167
- DARPA (U.S. Defense Advanced Research Projects Agency), 9–10
- FORCES project, 10
- data
 - external, application building, 167–168
 - insights from, 4–6
 - speed, 67–68
 - streaming, 67–68
- data access, 18
- cognitive systems, extraction, 27–28
- data analytics services, 18–19
- data catalogs, 18
- data integration, future of cognitive computing, 245
- data mining, advanced analytics and, 92–93
- Data Nodes, 65
- data ownership, 205–206
- data patterns, machine learning and, 29
- data services, 61
- data streaming, 67–68
- data visualization, 20
- data warehouses, 61–62
- databases
 - Big Data, 58–61
 - columnar, 60
 - document databases, 59–60
 - graph, 60
 - KVP (Key-Value Pair), 59
 - Polyglot Persistence, 60
 - PostGIS/OpenGEO, 60
 - spatial, 60
 - structured data, 59–61
 - unstructured data, 59–61
 - KVP (Key-Value Pair), 59
 - NoSQL, 61
 - SQL, 61
- decision tree, supervised learning and, 95
- DeepFace, 102
- DeepQA
 - future of cognitive computing, 243
 - Watson, 138, 142–144
 - architecture, 144–154
- DENDRAL project, 9
- descriptive analytics, 87–88
- deterministic system, 5
- DFNN (Dynamic Fuzzy Neural Network), 102
- dimensionality reduction, 62
- discourse analysis, NLP, 46–47
- disruptive models, 126–127
- distributed computing
 - cloud computing and, 111–112
 - shared resources and, 109–110
- distributed data management, 37
- distributed file system, 61
- document databases, 59–60
- domains for cognitive computing
 - use, 5–6
- DOR (Digital On-Ramps) project, 215–216
- E**
- EFLD (Eigenface-Fisher Linear Discriminant), 102
- EM-algorithm, 97
- energy management, government applications, 209–210
- ETL (extract-transform-load) tools, 26, 61
- expert systems, 9
- external data, application building, 167–168
- F**
- facial recognition, 101–102
- Feigenbaum, Edward, 9
- files, distributed filesystem, 61
- financial services, 232
- FORCES project, 10
- fraud detection, 53
- future applications, 239–249
 - best practices packaging, 238–239
 - human-to-machine interfaces, 237–238
 - knowledge management, 236–237
 - predictability, 236
 - technical advancements, 239
- G**
- GAAP (Generally Accepted Accounting Principles), 25
- games, AI and, 7
- GCA (Grid Cybersecurity Analytics), 211–212
- GenieMD, 191
- governance, cloud computing, 121–122
- government applications, 197–198
 - citizen-produced data, 202
 - city characteristics, 199–204
 - community infrastructure, 217–218
 - data integration, 203–204

government applications
 (*continued*)
 data ownership, 205–206
 data value, 205–206
 healthcare, 212–213
 Internet of Everything,
 204–205
 open data movement, 204
 system of systems, 199
 technology, 206–212
 transportation infrastructure,
 213–215
 workforce and, 215–217
 graph databases, 60

H

Hadoop, 63–67
 AM (Application Master), 66
 Avro, 66
 Cassandra, 66
 Chukwa, 66
 dark data, 68–69
 Data Nodes, 65
 HBase, 66
 Hive, 66
 Mahout, 66
 NameNodes, 64–65
 Pig, 66
 RM (Resource Manager), 66
 Spark, 67
 Tez, 67
 YARN (Yet Another
 Resource Negotiator), 66
 ZooKeeper, 67
 hardware architecture
 and future of cognitive
 computing, 245–249
 HBase, 66
 HDFS (Hadoop Distributed
 File System), 63–64
 healthcare application
 approaches, 181–182
 Big Data analytics, 180–181
 clinical teaching and, 193–195
 corpus, 184, 185
 data service providers, 179
 EMR (electronic medical
 record), 191–193
 foundations, 176–177
 government agencies, 179
 health improvement, 186–191
 ontologies, 182–183
 patients, 178
 patterns, 179–180
 payers, 179

pharmaceutical companies,
 178
 providers, 178
 questions, 183–185
 system training, 185
 healthcare in government
 applications, 212–213
 hierarchies, 73
 HMMs (Hidden Markov
 Models), 48–49
 horizontal scaling, 111
 human-generated data, 53–56
 hybrid cloud, 115–116
 hypotheses, 1, 33–34
 generation, 19, 34–35
 Watson (IBM), 152–153
 scoring, 35–36

I

IaaS (Infrastructure as a
 Service), cloud computing,
 117
 containers, 118
 SDE, 118
 virtualization, 117–118
 IBM
 grand challenges, 139
 Watson (*See* Watson (IBM))
 image analytics, 101–103
 images, analyzing, 16
 industry-specific taxonomies
 and ontologies, 168
 inferential statistics, 29
 infrastructure, 37
 Big Data, 58
 in-memory capabilities,
 104–105
 insight gained from data, 4–6
 intuition, 13–14
 IoT (Internet of Things), 4
 NLP and, 51

J

Jeopardy!, Watson and, 139–142
 judgment, 12–13

K

Kahneman, Daniel, 12–13
 KDE (kernel density
 estimation), unsupervised
 learning and, 97
 K-means algorithm, 97

k-NN (k-Nearest Neighbor),
 supervised learning and, 96
 knowledge, businesses and,
 127–128
 planning and, 131–133
 knowledge representation,
 71–73
 chess notation, 76–77
 classification, 75
 ignored information, 78–79
 implementation, 85
 models, 80–85
 multiple views, 79–80
 ontologies, 81–83
 definition, 73–75
 semantic web, 83–84
 simple trees, 83
 taxonomies, 80–81
 definition, 73–75
 KVP (Key-Value Pair), 59

L

labeled data, 94
 language
 lexical analysis, 45
 NLP and, 43
 law enforcement, government
 applications, 207–208
 learning, 1
 learning process, 19
 Lederberg, Joshua, 9
 legal applications of cognitive
 computing, 232
 lexers, lexical analysis and, 45
 lexical analysis, 45
 linguistic analysis, 40
 linguistics, 43
 construction grammars, 46
 discourse analysis, 46–47
 pragmatics, 47
 syntax, 45–46
 Logic Theorist, 8

M

machine learning, 4
 advanced analytics and,
 93–98
 analytics maturity and, 88
 customers and, 90
 data patterns, 29
 government applications,
 211–212
 inferential statistics, 29
 labeled data, 94
 reinforcement learning, 31–32

- SLA and, 91
 - supervised learning, 29–31, 94
 - decision tree, 95
 - k-NN, 96
 - neural networks, 95
 - regression, 95
 - SVM, 96
 - unsupervised learning,
 - 32–33, 96
 - clustering, 97
 - KDE, 97
 - NMF, 97
 - PCA, 97
 - SOM, 97–98
 - SVD, 97
 - management services, 18
 - MapReduce, 63–64
 - marketing applications for
 - cognitive computing, 232–233
 - markets for cognitive computing
 - characteristics, 222–223
 - logistics, 228–229
 - retail, 224–226
 - security, 230–231
 - telecommunications, 229–230
 - threat detection, 230–231
 - transportation, 228–229
 - travel, 226–227
 - Markov Models, 48–49
 - MARLIN-ATSC (Multi-Agent Reinforcement Learning Integrated Network of Adaptive traffic Signal Controllers), 214–215
 - maturity levels analytics, 88, 132–133
 - McCarthy, John, 8
 - Medical Sieve, 102
 - metadata, 18
 - military, DARPA, 9–10
 - Minsky, Marvin, 8, 78–79
 - models, 1, 21
 - supervised learning, 30
 - unsupervised learning, 32
 - morphology, 44
 - MSPs (Managed Service Providers), 114–115
 - multitenant cloud, 112–113
- N**
- NameNodes, 64–65
 - narrative solutions,
 - presentation, 36
 - neural networks, supervised learning and, 95
 - neurology, 11
 - neurosynaptic architectures, 246–248
 - Newell, Allen, 8
 - “Chess Playing Programs and the Problem of Complexity,” 8
 - NLP (Natural Language Processing), 4, 5, 10, 39–40
 - business problems, 50–53
 - construction grammars, 46
 - context, linguistic analysis, 40
 - customer’s voice and, 51–52
 - description, 6
 - disambiguation, 47–48
 - discourse analysis, 46–47
 - fraud detection and, 53
 - future of cognitive computing, 244–245
 - history, 43
 - IoT and, 51
 - language, identification, 43
 - lexical analysis, 45
 - linguistics, 43
 - Markov Models, 48–49
 - morphology, 44
 - phonology, 44
 - pragmatics, 47
 - RDF, 50
 - semantic web, 50
 - Sentiment Analysis, 52
 - shopping experience and, 50–51
 - syntax, 45–46
 - tokenization, 43
 - Watson and, 137–138
 - words, connecting, 42–43
 - WSD, 49
 - NMF (nonnegative matrix factorization),
 - unsupervised learning and, 97
 - NoSQL, databases, 61
 - NUMA (Non-Uniform Memory Access), 63
 - numerical linear algebra, 62
- O**
- ontologies, 25, 81–83
 - definition, 73–75
 - industry-specific, 168
 - open data movement,
 - government applications and, 204
 - open source tools, advanced analytics and, 105–106
- P**
- PaaS (Platform as a Service),
 - cloud computing and, 120
 - parallelism, 27
 - patterns
 - machine learning and, 29
 - unsupervised learning, 32
 - PCA (Principal Components Analysis)
 - images, 101–102
 - unsupervised learning and, 97
 - persistence, cognitive systems, 84–85
 - phonology, 43
 - physical foundation of Big Data, 58
 - Polyglot Persistence, 60
 - PostGIS/OpenGEO, 60
 - pragmatics, 47
 - predictive analytics, 88
 - business value, 98–99
 - use cases, 99
 - prescriptive analytics, 88
 - presentation services, 20, 36–37
 - private cloud, 114
 - probabilistic system, 5
 - proximity, patterns and, 29
 - psychology, 11
 - public cloud, 112–114
- Q**
- quantum architectures, 248–249
 - Question Analysis (Watson), 148–152
 - question-answer pairs in application building, 164–164
- R**
- R language, 105–106
 - RDBMS (relational database management system), 130
 - RDF (Resource Description Framework), 50, 83
 - regression, supervised learning and, 30–31, 95

reinforcement learning
 (machine learning), 31–32
 reporting services, 36
 representing knowledge, 71–73
 chess notation, 76–77
 classification, 75
 ignored information, 78–79
 implementation, 85
 models, 80–85
 multiple views, 79–80
 ontologies, 81–83
 definition, 73–75
 semantic web, 83–84
 simple trees, 83
 taxonomies, 80–81
 definition, 73–75
 rule-centric thinking, 14–15

S

SaaS (Software as a Service), 18
 cloud computing and,
 118–120
 Samuel, Arthur Lee, 7–8
 scaling, cloud computing and,
 111–112
 scoring, Watson (IBM) and,
 153–154
 SDE (Software Defined
 Environment), 118
 security
 cloud computing, 121–122
 government applications, 200
 infrastructure, Big Data, 58
 semantic web, 50, 83
 Sentiment Analysis, 52
 serialized service, 61
 shared resources, distributed
 computing, 109–110
 Shaw, Cliff, “Chess Playing
 Programs and the Problem
 of Complexity,” 8
 shopping, NLP and, 50–51
 Simon, Herbert A., 8
 “Chess Playing Programs
 and the Problem of
 Complexity,” 8
 sketching and streaming
 algorithms, 62
 SLA (service level agreements)
 analytics and, 91
 cloud computing and, 113
 machine learning and, 91
 Smart+Connected
 Communities Initiative,
 217–218

SOM (Self Organizing Map),
 unsupervised learning
 and, 97–98
 spatial databases, 60
 speech analytics, 103
 SQL (Structured Query
 Language), databases, 61
 state, cognitive systems, 74–75,
 84–85
 statistics
 advanced analytics and, 92
 inferential, 29
 structured data, 59–61
 supervised learning, 94–95
 decision tree, 95
 k-NN, 96
 machine learning, 29–31
 neural networks, 95
 regression, 95
 SVM, 96
 SVD (Singular Value
 Decomposition),
 unsupervised learning
 and, 97
 SVM (Support Vector
 Machine), supervised
 learning and, 96
 syntax, 45–46
 System 1 (Automatic
 Thinking), 13–14
 System 2 (Controlled, Rule-
 centric, Concentrated
 Effort), 14–15

T

taggers, lexical analysis and, 45
 taxonomies, 18, 80–81
 definition, 73–75
 industry-specific, 168
 ontologies and, 25
 technical advancements
 affecting future
 applications, 239
 testing, application building
 and, 171–173
 text analytics, 99–101
Thinking Fast and Slow
 (Kahneman), 13
 tokenization, NLP and, 43
 tokens, lexical analysis and, 45
 training, application building
 and, 171–173
 training tools, future of
 cognitive computing,
 244–245

transportation, government
 applications, 213–215
 trees, 83
 Turing, Alan, 6–7

U

UIMA (Unstructured
 Information Management
 Architecture), 143
 unstructured data, 59–61
 unsupervised learning, 96
 clustering, 97
 KDE, 97
 machine learning, 32–33
 NMF, 97
 PCA, 97
 SOM, 97–98
 SVD, 97

V

vertical scaling, 111
 video, analyzing, 16
 virtualization
 cloud computing and, 118
 IaaS (Infrastructure as a
 Service), 118
 visualization, 5
 visualization services, 20,
 36–37

W-Z

Watson (IBM)
 commercial applications,
 141–144
 DeepQA, 138, 142–144
 architecture, 144–154
 Jeopardy!, 139–142
 overview, 137–138
 software architecture,
 142–144
 Weiner, Norbert, 7
 Welltok, 187–191
 words
 connecting for meaning,
 42–43
 morphology, 44
 workflow services, 61
 WSD (Word-Sense
 Disambiguation), 49