

INDEX

A

- Adaptive resonance theory..... 199
- ADIT. *See* AutoDep Input Tool
- AFS. *See* Andrew file system
- Agglomerative information bottleneck
 - (AIB) 180–181
- Andrew file system (AFS) 14–15
- Annotation and image markup
 - (AIM) project 8–9
- Application programming
 - interfaces (API) 229, 231, 251
- Architectural extension, semantically based integration
 - automatic query generation algorithm..... 45–46
 - data pointer/URL concept..... 46
 - postprocessor 47
 - preprocessor 46–47
 - term-index-source 43–45
- Architecture and design, grid wizard
 - enterprise system
 - agent components..... 231, 232
 - application programming interfaces (API) 229
 - client components..... 231–232
 - daemon components..... 230–231
 - subsystem components 230
- Asymmetric unit (ASU) 161
- AutoDep Input Tool (ADIT)..... 160
- Automatic query generation algorithm..... 45–46

B

- Bayesian network..... 202–230
- Betweenness centrality 78
- Bioinformatic Sequence Markup
 - Language (BSML) 83
- Biological micro-compartment (BMC) fold 166–167
- Biomedical Informatics Research Network
 - (BIRN)
 - condor..... 250
 - curation and annotation..... 30
 - data integration..... 28
 - data management interface
 - curatorial environment..... 28
 - data submission process 28–29
 - data use and discovery 31
 - Grid Account Management Architecture
 - (GAMA) 249–250

- grid portal..... 242
- grid security infrastructure (GSI)
 - delegation 248–249
 - proxy credentials 247–248
 - public key cryptography..... 246
 - secure socket layer (SSL) 247
 - X.509 certificates 246–247
- Grid Wizard (GWiz) 250
- human subjects' data 29–30
- ImageJ 251–252
- Java Network Launch Protocol (JNLP) 251
- Java Web Start 251
- MyProxy server..... 249
- public key infrastructure (PKI) concepts
 - authentication and confidentiality 244–245
 - cryptography, symmetric and asymmetric
 - ciphers..... 243–244
 - data integrity and nonrepudiation..... 245–246
 - public-private key pair 244
 - X.509 certificates 242–243
- Storage Resource Broker (SRB) 250
- Biomedical organizations, data management
 - database
 - common user interfaces 5
 - data retrieval and storage 4
 - Google's search engine..... 5
 - thick clients..... 4
 - thin clients 4, 5
 - Wikipedia 5
 - data federations and data warehouses 5–6
 - information interchange languages
 - annotation and image markup
 - (AIM) project 8–9
 - clinical data interchange standards
 - committee (CDISC)..... 7–8
 - digital imaging and communications
 - in medicine (DICOM) 11–12
 - health level 7 (HL7) 10–11
 - XML-based clinical experiment data
 - exchange (XCEDE)..... 7
- BSML. *See* Bioinformatic Sequence Markup Language

C

- Cancer central clinical database (C3D) 20–21
- CDISC. *See* Clinical data interchange standards committee
- Ceph..... 17–18

CFT formalism. *See* Conformal field theory formalism
Clinical assessment layout manager (CALM)..... 19
Clinical data interchange standards committee
 (CDISC)..... 7–8
Closeness centrality..... 65
CMS. *See* Content management based systems
Coda..... 15–17
Concept mapping tool..... 30, 47–51
Conditional random fields (CRFs)..... 185–187
Conformal field theory (CFT) formalism
 super-gravity AdS and Yang–Mills..... 122
 tangent vector space..... 121
 toy 3-dim manifold..... 122–123
Content management based systems (CMS)..... 21–22
Cross-population extended haplotype
 homozogosity test..... 92
Cross-space mapping..... 209
Cytoscape..... 79

D

DARPA Agent Markup Language (DAML)..... 83
Database registration tool..... 47, 48
Data integration methods..... 28
 clique structure..... 71
 compound graph..... 72
 genetic interaction network..... 71
 pathDB system..... 71–72
 pathSys schema..... 73
 pathways database system..... 72
 phenotype classification tree..... 73
 querying graphs
 betweenness centrality..... 78
 BioNetSQL..... 76
 directed acyclic graph (DAG)..... 77
 extractNodes function..... 78
 gene ontology..... 77, 78
 pathexpr function..... 78
 relational-style query..... 76
 stat-source..... 78
 query tool..... 72
 schema mapping
 conflict resolution..... 75–76
 yeast phenotype classification..... 73, 74
 software tools, graph-based network
 analysis..... 78–79
 XML-RDF-OWL..... 73
Data pointer/URL concept..... 46
Data submission process..... 28–29
Davies–Bouldin (DB) index..... 180
Degree centrality..... 65, 66
Denial of Service (DoS) attacks..... 261
Digital imaging and communications in medicine
 (DICOM)..... 11–12
Digital signature..... 245–246

Directed acyclic graph (DAG)..... 77
Distinguished Name (DN) strings..... 247–248
Distributed biomedical data management systems
 cancer central clinical database (C3D)..... 20–21
 content management based
 systems (CMS)..... 21–22
 extensible neuroimaging archive
 toolkit (XNAT)..... 20
 human imaging database (HID)
 base tuple and clinical assessment
 layout manager (CALM)..... 19
 open source data management system..... 18
Distributed data storage
 Andrew file system (AFS)..... 14–15
 Ceph..... 17–18
 Coda..... 15–17
 storage resource broker (SRB)..... 13–14
Distributed heterogeneous data resources..... 40–41

E

End entity certificate (ECC)..... 250
Ensembl registry points..... 111
Enterprise system, user types
 administrators..... 225–227
 console user interfaces..... 227, 228
 end users tasks..... 227
 processes parallel execution
 language (P2EL)..... 227–229
Expectation maximization
 (EM) algorithm..... 175–177
Expression quantitative trait loci (eQTLs)..... 93
eXtensible Markup Language (XML)..... 79, 82–83
Extensible neuroimaging archive
 toolkit (XNAT)..... 20

F

Farthest-neighbor agglomerative
 hierarchical clustering..... 179
Federated biomedical data management
 environment
 distributed biomedical data management systems
 cancer central clinical database (C3D)..... 20–21
 content management based
 systems (CMS)..... 21–22
 extensible neuroimaging archive
 toolkit (XNAT)..... 20
 human imaging database (HID)..... 18–19
 distributed data storage
 Andrew file system (AFS)..... 14–15
 Ceph..... 17–18
 Coda..... 15–17
 storage resource broker (SRB)..... 13–14
Feedback and gene circuits..... 61

G

Gene mutation databases..... 99–101
 Gene ontology (GO)..... 66
 Gene regulation databases 97–99
 Genetic variation patterns, non-coding regions
 cis-and trans-regulatory interactions 90, 91
 integrated haplotype score (iHS) 92
 linkage disequilibrium 90
 long range haplotype (LRH) test..... 92
 transcription-mediating proteins 90
 Graphic user interface (GUI) 47, 48
 Graph-theoretic methods, biological data analysis
 motifs and clusters
 distance-based clusters..... 70
 higher-order biological structure 68
 highly connected clusters 69
 local graph alignment algorithm 68
 node functional homogeneity 69
 overlapping clusters 69–70
 probabilistic algorithm 68
 PSI-BLAST 68
 subgraphs 67
 network topology and graph analysis..... 64–65
 subgraphs, pathways and functional
 modules
 bottleneck nodes 66
 centrality 65–66
 color coding 67
 directed acyclic graph (DAG) 66
 fundamental biological concepts 65
 gene ontology (GO) 66
 graph-driven analysis strategy..... 65
 MAPK..... 67
 pathway redundancy 66
 Grid security infrastructure (GSI)
 delegation 248–249
 proxy credentials 247–248
 public key cryptography..... 246
 secure socket layer (SSL) 247
 X.509 certificates 246–247
 Grid wizard enterprise (GWE) system
 architecture and design
 agent components 231, 232
 application programming
 interfaces (API) 229
 client components..... 231–232
 daemon components 230–231
 subsystem components 230
 cluster anatomy..... 222
 distributed system..... 222–224
 3D Slicer..... 229
 enterprise system, user types
 administrators 225–227
 end users tasks 227

 processes parallel execution language
 (P2EL)..... 227–229
 features
 robustness 225
 scalability 224–225
 value added 224
 LDDMM, BIRN portal..... 229
 mailing lists..... 238
 project website 237
 requirements 222
 software development process..... 233–234
 issues management system..... 236
 open source community 234–235
 project management 235–236
 user communication 236–237

H

Health Insurance Portability and Accountability
 Act (HIPAA)..... 3–4
 Health Insurance Privacy and Portability
 Act (HIPPA) 240–241
 Health level 7 (HL7)..... 10–11
 Hierarchical clustering algorithm 177–181
 HID. *See* Human imaging database
 Hidden Markov model (HMM) 134, 186
 Hierarchy of Brain Network Languages..... 198
 High performance computing (HPC)..... 219
 HIPAA. *See* Health Insurance Portability
 and Accountability Act
 Holographic network processor and calculus
 Bayesian network..... 202–203
 circular convolution/de-convolution 203, 205
 delocalization 204, 205
 holographic map 204–206
 neurons, inter-connections 205, 206
 primary concept node 201–202
 prototypes 205–206
 Human imaging database (HID)
 base tuple and clinical assessment layout
 manager (CALM) 19
 open source data management system 18
 Human subjects' data..... 29–30
 Hypertext Transfer Protocol (HTTP) 255

I

Imagination blending process, ontological model
 concepts 207–209
 mapping relations 211
 mental space and object universal structure 209
 ontological dK-series and dK-graphs 210
 term-object-map..... 211
 Information extraction, PubMed abstracts
 conditional random fields (CRF)
 HMM..... 186

- Information extraction, PubMed abstracts (*Continued*)
 parameter estimation 186–187
 probability..... 185–186
 datasets and preprocessing..... 187–188
 features 188–190
 NER performance
 biological named entities 190–192
 second named-entity set 192–194
- Information integration framework
 distributed heterogeneous data resources..... 40–41
 mapping relations
 joinable relation 42
 ontology-map relation 41
 value-map relation 42
 ontologies 41
 optimizations 43
 query and view definition languages
 aggregation function 42–43
 query definition 42, 43
 registry information 42
- Information interchange languages
 annotation and image markup
 (AIM) project 8–9
 clinical data interchange standards committee
 (CDISC)..... 7–8
 digital imaging and communications in medicine
 (DICOM)..... 11–12
 health level 7 (HL7) 10–11
 XML-based clinical experiment data exchange
 (XCEDE) 7
- Integrated development environment (IDE)..... 235
Integrated haplotype score (iHS) 92
InterProSurf 163
- J**
- Java Network Launch Protocol (JNLP) 251
Java virtual machine (JVM)..... 163
Jensen-Shannon (JS) divergence 181
Joint Center for Structural Genomics (JCSG)..... 158
- K**
- Knowledge discovery, neurodegenerative disease
 information extraction, PubMed
 biological named entities 190–192
 conditional random fields (CRF)..... 185–187
 datasets and preprocessing..... 187–188
 features..... 188–190
 second named-entity set 192–194
 MEDLINE 174, 195
 semi-supervised machine learning
 F_1 values..... 183–185
 meta-search engine 182–183, 195
 mixture component/topic number
 selection 177–181
 relevance feedback 174–175
 simulated annealing EM..... 181–182
 transductive SVM..... 175
 unigrams expectation maximization
 approach 175–177
- L**
- Large Deformation Diffeomorphic Metric
 Mapping (LDDMM) 229, 254
Long range haplotype (LRH) test..... 92
- M**
- Mapping relations, integration framework
 joinable relation 42
 ontology-map relation 41
 value-map relation 42
Maximum a posteriori (MAP) point
 estimation 176
Multiple mixture components per class
 EM (MMEM)..... 177
MyProxy server..... 249
- N**
- Naive Bayes classifier..... 175–176
Named entity recognition (NER)..... 185, 186
National Institute of General Medical
 Sciences (NIGMS) 158
Natural language processing (NLP) 185
- O**
- Object relational mapping (ORM)..... 230
Ontology web language (OWL) 215
- P**
- Parallel execution process
 cluster computing 220
 grid wizard enterprise (GWE)
 architecture and design 229–232
 cluster anatomy 222
 distributed system 222–224
 3D Slicer..... 229
 enterprise system, user types 225–229
 features..... 224–225
 LDDMM, BIRN portal..... 229
 requirements 222
 mailing lists..... 238
 parallelization problem 220–221
 project website 237
 software development process..... 233–234
 issues management system..... 236
 open source community 234–235
 project management 235–236
 user communication..... 236–237

Part-of-speech (POS) tags	188
PathSys schema	62–63, 73
Pathway redundancy.....	66
Pathways database system.....	62
Personal Health Information (PHI).....	240
Pfaff dimension	125
Planner, system architecture	
disjunctive normal form (DNF)	38
heuristics.....	36–39
mediator query.....	36
unfolding, query rewriting procedure	37–38
well behaved plan.....	38
Postprocessor.....	47
Preprocessor.....	46–47
Protein interactions calculator (PIC).....	163
Protein Interfaces, Surface and Assemblies	
(PISA) Service.....	161–162
Protein-protein interaction (PPI) server.....	163
Protein quaternary structure (PQS) server	161
Protein structure determination	
biomolecules, protein-protein interface	
crystallographic and noncrystallographic	
information	161
FastContact	163
interaction analysis.....	164
PIC and PPI servers	163
PISA server	161–162
ProFace.....	163
SPPIDER.....	163, 164
domain architecture	131–132
domain assignment	150–151
domain duplication.....	150
3-D structural homolog search	168–169
biological micro-compartment	
(BMC) fold.....	166–167
DALI Server.....	165–166
MATRAS and FATCAT	167
target database	166
TOPSCAN algorithm.....	165–166
meta-servers and integrated prediction	
services.....	143–145
model building	
model refinement.....	141
side-chain conformations.....	140
software	141, 142, 153
unaligned regions.....	140–141, 152
model evaluation and	
improvement.....	141–144, 152–153
procedure	
domain boundaries and folds.....	146–147
3D protein model	148, 152–153
model evaluation.....	148–149, 152
objective.....	145–146
sequence-template alignments.....	148, 152
template model	147, 149, 151–152
Protein Structure Initiative Structural	
Genomics centers.....	158
schematic representaion.....	131–132
sequence similarity detection	
domain boundaries and folds.....	132–134, 149–151
fold assignment, evaluation	
and improvement.....	137–139, 151
local structural feature prediction	
method.....	136–137
methods	134–137, 152
sequence alignment, evaluation	
and improvement.....	139
validation	
molprobability and NCS check	160
quality control (QC) Script and check.....	159–161
refinement statistics and PDB check.....	159
resolve check.....	160
sequence check.....	159
web servers and programs.....	144, 146
Protein Structure Initiative Structural	
Genomics centers.....	158
Public data sharing	
BIRN data repository (BDR)	
curation and annotation.....	30
data management interface	28–29
data use and discovery.....	31
human subjects' data	29–30
federated informatics infrastructure.....	26
Public key infrastructure (PKI) concepts	
cryptography, symmetric and asymmetric	
ciphers.....	243–244
data integrity and nonrepudiation	245–246
public-private key pair	244
X.509 certificates	242–243
PubMed search engine	182, 187, 195
Q	
Query Atlas meta-search engine	182, 194, 195
R	
Regulatory polymorphisms	
genetic variation patterns, non-coding regions	
cis- and trans-regulatory interactions	90, 91
integrated haplotype score (iHS).....	92
linkage disequilibrium	90
long range haplotype (LRH) test.....	92
transcription-mediating proteins	90
materials	
derived alleles and frequency	101–102
gene regulatory annotation resources	97–99
single nucleotide polymorphism	
resources	99–101
transcription factor binding sites	96–97
methods	
ancestral and derived allele frequency	107–109

- Regulatory polymorphisms (*Continued*)
- ORegAnno 104–105
 - PupaSuite, regulatory variation analysis 110
 - single nucleotide polymorphisms 105–107
 - Sockeye, regulatory variation analysis 110
 - transcription factor binding sites and regulatory variants, relationships 109, 110
 - transcription factor binding sites, weight matrices 102–104
 - prioritization, computation methods
 - allele-specific expression 95
 - electrophoretic mobility shift 93
 - functional regulatory variation 95
 - ORegAnno and PAZAR 94
 - PromoLign 93
 - purple sea urchin 95
 - regulatory mutations 94
 - selection pressure 95
 - transcription start sites (TSS) 93
 - role and extent, humans
 - allele-specific gene expression 92, 93
 - Marfan syndrome 93
 - phenotypic diversity and regulatory variation 92
 - scalability, expression technology 93
 - specific mutations 92
 - Relevance feedback 174–175
 - Resource Description Framework (RDF) 82, 83
 - Ricci flow 123, 124
- S**
- SCOP database
- ASTRAL resource 135
 - libraries, protein domains 133
 - protein sequence similarity 138
- Semantic data integration environment
- architectural extensions
 - automatic query generation algorithm 45–46
 - data pointer/URL concept 46
 - postprocessor 47
 - preprocessor 46–47
 - term-index-source 43–45
 - integration framework
 - distributed heterogeneous data resources 40–41
 - mapping relations 41–42
 - ontologies 41
 - optimizations 43
 - query and view definition languages 42–43
 - mediator clients and user interaction
 - concept mapping tool 47–51
 - database registration tool 47, 48
 - webservice API 50–51
 - system architecture
 - executor 39
 - gateway 35–36
 - planner 36–39
 - registry 36
 - wrapper service 39–40
- Semi-supervised machine learning
- F_1 values 183–185
 - meta-search engine 182–183, 195
 - mixture component/topic number selection
 - agglomerative information bottleneck (AIB) 180–181
 - cluster validity indices 179–180
 - cross-validation 177, 178
 - hierarchical clustering algorithm 177, 179
 - relevance feedback 174–175
 - simulated annealing EM 181–182
 - transductive SVM 175
 - unigrams expectation maximization (EM) approach
 - multi topic/mixture component per class model 177
 - Naive Bayes classifier 175–176
- Sequence similarity detection, protein structure
- domain boundaries and folds
 - algorithms, protein domain prediction 134
 - annotated protein domain database 133
 - domain duplication and assignment 150
 - protein domain lengths 151
 - query sequence 132–133
 - fold assignment
 - consistency 138, 151
 - intermediate sequence similarity searches 138
 - quality evaluation, sequence profiles 138–139
 - local structural feature prediction 136–137
 - methods
 - alignment accuracy 135–137, 152
 - fold prediction 134–135
 - sequence alignment determination 139
- Simulated annealing expectation maximization (SAEM) 183
- Single nucleotide polymorphisms
- gene mutation databases 99–100
 - HapMap data, genotype data 107
 - rs numbers 105, 106, 111
 - ss numbers 111
 - UCSC Genome Browser, genetic variation data 106–107
- Single Sign-On (SSO)
- analysis and visualization
 - BIRN digital certificate 256, 259
 - ImageJ application, local machine 257, 260
 - information retrieval 255–256
 - username and proxy, GSI host 256, 258
 - BIRN technology

condor.....	250
GAMA.....	249–250
grid portal.....	242
grid security infrastructure (GSI).....	246–249
Grid Wizard (GWiz).....	250
ImageJ.....	251–252
JNLP and Java Web Start.....	251
MyProxy server.....	249
public key infrastructure (PKI)	
concepts.....	242–246
Storage Resource Broker (SRB).....	250
DOS attacks and database security.....	261
discussion.....	239–240
grid storage access	
motivation.....	252
process.....	252–254
motivations	
HIPPA requirements.....	240–241, 259–260
software security.....	241–242
remote jobs submission	
grid portal.....	255
LDDMM.....	254–255
reviewing and monitoring.....	256
script attacks, blogs and forums.....	261
SQL injection.....	260
vulnerability, Java Web Start (JWS).....	261
Software development, grid wizard enterprise system	
development process.....	233
issues management system, Jira.....	236
open source community	
development tools.....	234
product libraries.....	235
project management.....	235–236
project-related activities.....	233–234
user communication.....	236–237
Solvent Accessibility Based Protein–Protein Interface	
iDentification and Recognition	
(SPPIDER).....	163
Storage resource broker (SRB).....	13–14, 250
Structural alignment servers. <i>See</i> Structural	
homolog search	
Structural Genomics (SG).....	157
Structural homolog search.....	168–169
biological micro-compartment	
(BMC) fold.....	166–167
DALI Server.....	165–166
MATRAS and FATCAT.....	167
target database.....	166
TOPSCAN algorithm.....	165–166
Support vector machines (SVMs).....	174
Symmetric Kullback–Leibler (KL) divergence.....	179
System architecture	
executor.....	39
gateway.....	35–36
general architecture, data integration system.....	35
planner	
disjunctive normal form (DNF).....	38
heuristics.....	36–39
mediator query.....	36
unfolding, query rewriting procedure.....	37–38
well behaved plan.....	38
registry36	
wrapper service.....	39–40
System biology	
auto poetic functionality, coherent structure.....	125
CFT formalism	
super-gravity AdS and Yang–Mills.....	122
tangent vector space.....	121
toy 3-dim manifold.....	122–123
data integration methods	
clique structure.....	71
compound graph.....	72
genetic interaction network.....	71
pathDB system.....	71–72
pathSys schema.....	73
pathways database system.....	72
phenotype classification tree.....	73
querying graphs.....	76–78
query tool.....	72
schema mapping.....	73–76
software tools, graph-based	
network analysis.....	78–79
XML-RDF-OWL.....	73
delocalization.....	116, 117
dynamic bio-system analysis.....	118–119
dynamic core	
functional cluster.....	120
information integration	
and causal power.....	119, 120
gene expression and gene regulation	
exons and introns, splicing.....	60
feedback and gene circuits.....	61
post-translational modification.....	60–61
transcriptional modules.....	58–59
translation rate.....	60
geometric flow, delocalization and function	
conservation	
holographic renormalization group flow.....	123
R-curvature.....	124
Ricci flow.....	123, 124
Riemann manifold.....	124
graph-theoretic methods	
motifs and clusters.....	67–70
network topology and graph analysis.....	64–65
subgraphs, pathways and	
functional modules.....	65–67
holographic principle conjuncture.....	117
holographic renormalization group flow.....	123

- System biology (*Continued*)
informational manifold 118
KEGG 62
ontologies
BioEntity 80
Bioinformatic Sequence Markup
Language (BSML) 83
chemical hierarchy 80
DARPA Agent Markup Language
(DAML) 83
eXtensible Markup Language
(XML) 79, 82–83
gene classification 81–82
gene ontology 81
namespaces 83–84
Resource Description Framework
(RDF) 82, 83
paradigm 63
shortest path functionality 117
smart-node architecture, communication
nodes 116
statistical manifold (AdS formalism) 121
system state 118
- T**
- Term-index-source 43–45
Term-object-map 211
Text animation brain model
artificial intelligence (AI) and imagination process
Adaptive Resonance Theory 199
causal complexity 198
Hierarchy of Brain Network Languages 198
higher-level mental functions 200
memory concept 199
neural network dynamics optimization 198
two-sided blending 197
dK-ontological confabulation 213–215
graph calculus and generic space 212–213
holographic network processor and calculus
Bayesian network 202–203
circular convolution/de-convolution 203, 205
delocalization 204, 205
holographic map 204–206
neurons, inter-connections 205, 206
primary concept node 201–202
prototypes 205–206
ontological model, imagination
blending process
concepts 207–209
mapping relations 211
mental space and object universal
structure 209
ontological dK-series and dK-graphs 210
term-object-map 211
renormalization, dimension reduction 213
ScriptWriter and animation process
application 216–217
Ontology Web Language (OWL) 215
screen shot 216
text filtering 211–212
Transcription factor binding sites
prediction
TRANSFAC and JASPAR 96, 97
weight matrices 102–104
regulatory variants, relationships 109
Transport Layer Security (TLS) 247
- U**
- Unigrams expectation maximization mixture
maximum a posteriori (MAP)
point estimation 176
multi topic/mixture component
per class model 177
Naive Bayes classifier 175–176
- V**
- Velocity template language (VLT) 227
- X**
- XML-based clinical experiment data
exchange (XCEDE) 7
XNAT. *See* Extensible neuroimaging archive toolkit