
Index

- abbreviation expansion, 163
- ABGene, 157
- activator genes, 182
- acute lymphoblastic leukemia, 184
- agents, 94
 - cooperative, 102
 - multi-agent, 101
 - multi-agent systems, 200
- Ames test, 25
- artificial neural network, 22, 23, 28, 181, 191

- backward elimination, 23, 24
- bag-of-words model, 158
- Bayesian network, 7, 185, 189
- Bayesian score, 185
- biased competition, 199
- bibliometrics, 151
- BioCreative
 - concept identification vs. entity identification, 155
 - corpus, 157
 - standardization of the entity identification task, 154
- bioinformatics, 95, 131
- biological imaging, 198
 - fMRI, 197
- biological network, 179, 199
- boolean networks, 182

- cancer, 51
 - classification, 51
 - subtypes, 51
- carcinogenicity, 20
- case, 9

- case-based reasoning, 12
- cell, 43
 - development, 46
 - extra-cellular space, 197
 - nerve cell, 43
 - proteome, 191
- central nervous system, 37, 43
- chemical ontology, 3
- Chimaera, 138
- Class Similarity Average criterion, 11
- classification, 51, 89
 - cancer, 51
- clustering, 38, 110
 - hierarchical, 111
 - k-means, 111
 - model-based, 38
- content words, 165
- controlled vocabulary, 134
- correlation coefficient, 23
- cross-validation, 57
- curse of dimensionality, 51, 53

- DAG-Edit, 139
- DAML+OIL, 135
- data mining, 31, 109, 133, 182, 191
- data model, 134
- data pre-processing, 43
 - discretization, 43
- description logics, 135
- development, 43
 - of rat, 43
- Direct Memory Access Parsing, 149
- directed acyclic graph, 185
- dissimilarity measure, *see* distance metric

- distance metric, 109
 - correlation distance, 110
 - Euclidean distance, 110
 - Minkowski distance, 110
- distributed artificial intelligence, 200
- DMAP, 149
- DNA, 25
 - damage, 25
- DNA microarrays, 179
- drug design, 31
- dynamic programming, 42

- eigenvalue, 47
- eigenvector, 47
- entity identification, 154–158
 - as a source of error in information extraction, 159
 - definition, 154
 - dictionary-based, 155
 - learning-based, 155
 - problems in bibliometrics, 152
 - rule-based, 154
- Entrez
 - for NLP, 164
- entropy, 41
- equivalence class, 185
- evolutionary algorithm, 54, 93, 95, 98
 - multi-objective, 61
- Extensible Markup Language, 134

- F-measure, 153
- feature selection, 23, 52, 54
- fitness function, 56
- FlyBase
 - as resource for NLP, 156
- forward selection, 23, 24
- function words, 165

- GAPSCORE, 156, 157
- gene, 25
 - annotation, 47, 49
 - co-regulation, 37
 - mutant gene, 25
 - network, 179
 - regulation, 37, 179
- gene expression, 37, 49, 51, 177, 191
 - artificial, 189
 - co-expression, 188
 - data, 43
 - profiles, 109
 - profiling, 51
 - temporal, 37, 39
 - time series, 41, 43
- Gene Ontology, 131, 166
 - annotation of expression arrays, 152
 - NLP, resource for, 163
- gene regulation, 178
- generative inverse modeling, 189
- genetic algorithm, 26, 54
 - fitness, 54
- genetic algorithms, 23, 24, 53
 - multiple objective, 56
- genetic networks, 179, 188, 191
- genetic programming, 93
- GENIA corpus
 - described, 157
- GENIES, 164
- genomics, 51
 - functional, 51
- genotoxicity, 19, 25
- good predictor set, 53

- harmonic mean, 153
- hidden Markov model, 37, 39, 49
- hierarchical clustering, 183
- HUGO
 - as resource for NLP, 156
- hyphenation
 - indicating negation, 161
 - tokenization, 160
 - variability in gene names, 155

- indexing, 165
- inductive logic programming, 8
- information extraction, 158–164
 - database curation and population, 152
 - evaluation, 163
 - learning-based, 159
 - rule-based, 158
- information retrieval, 164–167
 - categorization, 166
 - clustering, 166
 - content words, 165
 - definition, 164
 - indexing, 165
- inhibitor genes, 182
- integrate-and-fire, 201
- Intelligent Systems for Molecular Biology

- as venue for NLP work, 147
- k-nearest neighbor, 60
- kappa statistic, 153
- KeX, 157
- knowledge base, 132, 134
- knowledge discovery, 2, 132, 133
- lazy learning, 9
- learning, 95
 - Hebbian, 204
 - lazy, 12
 - statistical, 176
- leave-one-out cross-validation (LOOCV), 52
- lexical variant generation, 159
- LID, 12
- LocusLink
 - as a resource for NLP, 156
 - entries as output of concept-based NLP system, 149
- lvg (lexical variant generation), 159
- lymphoma, 60
- machine learning, 1, 2, 7, 53, 158, 176, 179, 188, 191
 - concept learning, 37
- Markov chain, 38, 43, 44, 47, 48
 - hidden, 38
- mass spectrometry, 90
- maximum likelihood, 42
- mean-field approach, 203
- Medical Subject Headings, 166
- MedMiner, 152
- MeSH, 166
- messenger-RNA, 176
- metaheuristics, 89
- MetaMap, 159
- metric, 43
 - distance, 43
- microarray, 37, 53, 131, 191
- molecular descriptor, 19, 20
 - charge-distribution-related, 21
 - constitutional, 20
 - geometrical, 21
 - quantum-chemical, 21
 - topological, 20
- morphology
 - linguistic, 148
- mRNA, 52
- multi-linear regression, 22
- mutagen, 25
- mutagenicity, 20, 25
- mutation, 25
- named entity recognition, *see* entity identification
- natural language processing, *see* NLP
- neural networks, 200
- neurodynamics, 199
- neuroscience, 198
- neurotransmitter, 43, 197
 - receptor protein, 43
- NLP
 - and AI, 149
 - bibliometric approaches, 151
 - and database population, 151
 - definition, 147
 - evaluation
 - bake-offs, 154
 - entity identification, 157
 - information extraction, 163
 - kappa statistic, 153
 - metrics, 153
 - test suite generation, 158
 - learning-based
 - abbreviation expansion, 163
 - entity identification, 155
 - information extraction, 159
 - macromolecular sequence analysis, 148
 - review papers, 167
 - rule-based
 - abbreviation expansion, 163
 - entity identification, 154
 - information extraction, 158
 - sentence tokenization, 160
- NP-hard, 89, 100, 102
- OilEd, 138
- ontology, 129
 - engineering, 130
- Open Biological Ontologies, 131
- optimization, 89
- overfitting, 63
- OWL, 135
- Pacific Symposium on Biocomputing
 - as venue for NLP work, 147
- part-of-speech tagging, *see* POS tagging
- partial least squares, 22, 23

- peptide, 46, 92
 - mass fingerprint, 90
- peptides, 87
 - mass spectrometry, 87
- POS tagging
 - as part of information extraction system, 161
 - by ABGene, 157
 - defined, 161
- precision
 - definition, 153
- predictive toxicology, 2
- Predictive Toxicology Challenge, 2
- principal component analysis, 23
- probability, 42
 - estimation, 42
- probability distribution, 39, 60
- PROMPT, 137
- PROPER, 157
- propositional representation, 2
- Protege, 137
- protein, 87, 89
 - annotation, 95
 - expression, 87
 - differential, 90
 - family, 95
 - folding, 95, 100
 - identification, 89, 90
 - interaction, 113
 - sequence analysis, 89
 - structural motifs, 100
 - structure prediction, 95
- proteome, 177
- proteomics, 87, 90, 94
- PubMed
 - for NLP, 164
- PubMed Central collection, 164

- quantitative structure-activity relationship, 19, 20
- quantitative structure-property relationship, 19, 20

- reaction kinetics, 180
- reaction-diffusion models, 180
- reasoning, 133
 - ontology, 134
 - probabilistic, 7
- recall
 - definition, 153
- relational representation, 3

- SAMBO, 139
- sampling, 57
 - stochastic universal, 57
- Schank, Roger, 149
- score-based learning, 185
- search, 44, 54, 89, 93, 102
 - global, 98
 - hill-climbing, 98
 - randomized, 44
 - space, 89
 - stochastic, 93
- sequence, 44
 - alignment, 44
 - multiple alignment, 94, 95
- Shaud, 9
- similarity measure, *see* distance metric
- similitude term, 12
- simulated annealing, 23
- simulation, 26, 186, 201, 203, 205
- spiking neurons, 201
- stem
 - in NLP, definition, 161
- stemming, 161
- string design, 55
- structure learning, 185
- structure-activity relationships, 2
- syntax
 - of natural language, 148
- system, v
- systems biology
 - indexing document collections, 165
- Systems Biology Markup Language, 134

- tagset, 161
- taxonomy, 134
- text data mining, *see* NLP
- thesaurus, 134
- time series, 38
- tokenization
 - definition, 157
 - sentence-level, 157, 160
 - word-level, 160
- toxicity, 20
- transcription factor, 178
- Tsujii, Jun'ichi, 155

- UMLS, 163