

Index

AGFL, 90
EAG, 88
1-pushdown
 automaton, 74
IPDA, 74
1VWG, 76
2-pushdown
 automaton, 74
2PDA, 74

abstract
 grammar, 108
 syntax, 108
active
 feature, 37
additive
 learning, 41
affix, 13
 ambiguity, 101
 derived, 87
 expression, 94
 inherited, 87
 rule, 93
agreement, 84
algol 60, 75
algol 68, 76
algorithm
 Perceptron, 41
 Rocchio, 38
 SBC, 39
 Winnow, 40
alternative, 76
AMAZON, 66
ambiguity
 affix, 101
 lexical, 14, 24, 25
 structural, 24
ambiguous, 78
 grammar, 78
 inherently, 79

 sentence, 78
analysis, 77
 syntax, 66
anaphora, 22
 resolution, 22
antonym, 15
application
 phase, 34
arity, 94
attachment
 PP, 25
 premodifier, 24
authorship attribution, 33
automaton
 1-pushdown, 74
 2-pushdown, 74
 nondeterministic, 78
average
 macro, 35
 micro, 35
 Precision, 49

balanced
 Winnow, 41
binary
 classification, 35
 frame, 67
 weighting, 42
BNF, 75
Bottom-Up
 parsing, 79

call, 94
card
 index, 5
categorization
 text, 31
CF
 grammar, 75
 transducer, 106
CFG, 75

choice
 finite, 74
Chomsky
 hierarchy, 74
 normal form, 77
class
 inference, 32
 profile, 32
classes
 large, 35
 small, 35
classification
 binary, 35
 content, 32
 document, 31
 mono, 34
 multi, 34
 simple Bayesian, 39
classifier, 31
 linear, 38, 69
classifiers
 example-based, 34
 linear, 33
 rule-based, 33
clause
 relative, 20
 verbal, 27
clustering
 term, 52
CNF, 77
co-occurrence, 8
collocation, 10, 17, 24
complement, 27
 elision, 28
 preposition, 27
compositional, 105
 transduction, 66
compound
 word, 12
concept, 70
concordance, 5
concrete
 syntax, 108
confluence, 78
conjunction, 7
consistent
 substitution, 84
content

 classification, 32
contents, 3
context
 free, 74
 sensitive, 74
Context-Free
 Grammar, 75
controlled
 language, 4
coordination, 20
corpus, 10, 34
 Reuters-22713, 36
 Reuters-8119, 36
cosine
 normalization, 38
curve
 learning, 34

database, 5
deep
 linguistics, 64
definition, 94
demotion
 parameter, 40
density
 distributions, 53
depassivation, 29
derivation, 76
 left-most, 78, 80
 right-most, 78, 80
derivational
 morphology, 12
derived
 affix, 87
determiner, 20
diagram
 precision/Recall, 48
difference
 set, 90
dimensionality, 36
direct
 production, 74
direction, 87
directly
 produce, 73
discourse
 structure, 19
disjunction, 7

- distributions
 - density, 53
- document, 5
 - classification, 31
 - processing, 3
 - profile, 32
 - seen, 35
 - text, 5
 - unseen, 35
- domain, 16, 94
- DORO
 - project, 64
- elision
 - complement, 28
- Email
 - filtering, 31
- epoch, 34
- equivalent
 - strongly, 76
 - weakly, 76
- error
 - rate, 34
- example
 - negative, 41
 - positive, 41
- example-based
 - classifiers, 34
- expansion, 71
 - massive, 72
 - query, 9
 - selective, 71
- expression
 - affix, 94
 - regular, 7
- extraction
 - term, 33
- extractor
 - NP, 25
- feature, 33
 - active, 37
 - selection, 50
- feature set
 - sparse, 36
- Fibonacci
 - grammar, 79, 106, 107
 - transducer, 107
- filtering

- Email, 31
- finite
 - choice, 74
- flat
 - lattice, 97
- folklore
 - law, 37
- forest
 - parse, 100
- form
 - sentential, 74
- frame, 23, 66, 67
 - binary, 67
 - nested, 68
 - NP, 24
 - phrase, 23
 - semantic, 108
 - syntax, 68
 - VP, 30
- free
 - context, 74
- frequency
 - term, 50
- full-text
 - retrieval, 5
- function
 - retrieval, 6, 7
 - similarity, 17
 - transfer, 105
- fuzzy
 - matching, 17
- generic
 - rules, 95
- GNF, 77
- Grammar
 - Context-Free, 75
- grammar, 10
 - abstract, 108
 - ambiguous, 78
 - CF, 75
 - Fibonacci, 79, 106, 107
 - inference, 76
 - input, 105, 110
 - MEKO, 97, 123
 - output, 105, 110
 - partition, 79, 117
 - simultaneous, 105

- W, 88
- graph
 - PR, 48
- Greibach
 - normal form, 77
- group, 94
- growth
 - vocabulary, 37
- guard, 90, 93, 94
- GWB, 100
- head, 20, 67, 94
- headedness, 67
- heuristic
 - Winnow, 42
- hierarchy
 - Chomsky, 74
- holonym, 15
- homonym, 14
- hypermym, 15
- hyponym, 15
- hypothesis
 - KeyPhrase, 19
 - KeyWord, 7
- image
 - processing, 3
- index, 106
 - card, 5
 - KWIC, 5
- indirect
 - object, 27
- inference
 - class, 32
 - grammar, 76
- inflexional
 - morphology, 12
- information
 - noisy, 50
 - retrieval, 3, 5
- information industry, 3
- inherently
 - ambiguous, 79
- inherited
 - affix, 87
- initial
 - symbol, 73
- input
 - grammar, 105, 110

- inversion, 107
- invertible
 - transducer, 110
- is-a
 - relation, 15
- issue
 - normalization, 36
- item, 5
- join, 88
- kernel, 27
- key, 5
- KeyPhrase
 - hypothesis, 19
- keyPhrase, 19
- KeyWord
 - hypothesis, 7
- keyword
 - multiple, 7
 - single, 7
 - weighting, 7
- KSTEM
 - stemmer, 13
- KWIC
 - index, 5
- language, 73, 76
 - controlled, 4
- large
 - classes, 35
- lattice, 96
 - flat, 97
 - powerset, 97
- law
 - folklore, 37
- LCS, 63
- learning
 - additive, 41
 - curve, 34
 - supervised, 33
- left-hand
 - side, 73, 94
- left-most
 - derivation, 78, 80
- lemma, 10, 11
- lemmatisation, 10
- lemmatizer, 24
- lexical

- ambiguity, 14, 24, 25
- variation, 8
- lexico-semantic
 - relation, 15
- lexicon, 84
- linear
 - classifier, 38, 69
 - classifiers, 33
 - weighting, 42
- linguistic
 - resources, 66
 - technology, 3
 - terminology, 10
 - variation, 8
- Linguistic Classification System, 63
- linguistics
 - deep, 64
 - shallow, 64
- macro
 - average, 35
- mail
 - routing, 31
- Management
 - Work-Flow, 31
- massive
 - expansion, 72
- matcher, 66
- matching
 - fuzzy, 17
 - semantic, 69
 - syntactic, 69
- measure
 - quality, 6
- MEKO
 - grammar, 97, 123
- member, 76, 94
- meronym, 15
- micro
 - average, 35
- modifier, 67
 - post, 20
 - pre, 20
- monitoring
 - news, 32
- mono
 - classification, 34
- morphological

- variation, 8
- morphology
 - derivational, 12
 - inflectional, 12
- multi
 - classification, 34
- multiple
 - keyword, 7
- name, 94
- narrowcasting, 32
- near
 - synonym, 15
- negative
 - example, 41
- nested
 - frame, 68
- news
 - monitoring, 32
- NLP, 9
- noisy
 - information, 50
- nominalization, 29
- nondeterministic
 - automaton, 78
 - split, 88
 - transducer, 105
- nonterminal
 - symbol, 73, 94
- normal form
 - Chomsky, 77
 - Greibach, 77
- normalization, 9, 110
 - cosine, 38
 - issue, 36
 - word-order, 29
- noun
 - phrase, 19
 - proper, 14
- NP, 19, 20
 - extractor, 25
 - frame, 24
 - similarity, 24
- object, 27
 - indirect, 27
- OCR, 64
- ontology, 15
- operation, 87

- operator
 - trec, 108
 - union, 94
- option, 94
- order
 - word, 28
- output
 - grammar, 105, 110
- overtraining, 35
- paradigm, 12
- parameter, 94
 - demotion, 40
 - promotion, 40
- parse
 - forest, 100
 - trec, 77
- parser, 10, 66, 77
- parsing, 10
 - Bottom-Up, 79
 - Top-Down, 80
- part
 - transduction, 106
- part-of
 - relation, 15
- part-of-speech, 10
 - tagging, 10
- partition
 - grammar, 79, 117
- Perceptron
 - algorithm, 41
- performance, 6
- periphrastic
 - synonym, 17
- phase
 - application, 34
 - training, 34
- phrase, 19, 66
 - frame, 23
 - noun, 19
 - preposition, 20, 25
 - verb, 27
- Phrase Structure Grammar, 73
- polysemy, 14
- Porter
 - stemmer, 13
- POS, 10
- positive

- example, 41
- Winnow, 41
- post
 - modifier, 20
- powerset
 - lattice, 97
- PP, 20, 25
 - attachment, 25
- PR
 - graph, 48
- pre
 - modifier, 20
- pre-processing, 8
- Precision
 - average, 49
- precision, 6
- precision/Recall
 - diagram, 48
- premodifier
 - attachment, 24
- preposition
 - complement, 27
 - phrase, 20, 25
- processing
 - document, 3
 - image, 3
 - text, 3
- produce
 - directly, 73
- production, 74
 - direct, 74
 - rule, 73
 - terminal, 74
- profile, 32
 - class, 32
 - document, 32
- project
 - DORO, 64
 - TOSCA, 89
- prolog, 110
- promotion
 - parameter, 40
- proper
 - noun, 14
- PSG, 73
- quality
 - measure, 6

- query, 5
 - expansion, 9
- range
 - threshold, 42
- ranking, 7, 48
- rate
 - error, 34
- recall, 6
- reduction, 79
- regular, 74
 - expression, 7
- relation
 - is-a, 15
 - lexico-semantical, 15
 - part-of, 15
- relative
 - clause, 20
- resolution
 - anaphora, 22
- resources
 - linguistic, 66
- restriction, 96
- retrieval
 - full-text, 5
 - function, 6, 7
 - information, 3, 5
- Reuters-22713
 - corpus, 36
- Reuters-8119
 - corpus, 36
- right-hand
 - side, 73, 94
- right-most
 - derivation, 78, 80
- Rocchio
 - algorithm, 38
- root, 73, 93
- routing
 - mail, 31
- rule, 75, 76, 94
 - affix, 93
 - production, 73
 - schema, 84
 - syntax, 94
- rule-based
 - classifiers, 33
- rules
 - generic, 95
- SBC
 - algorithm, 39
- schema
 - rule, 84
- search
 - single-key, 7
 - string, 8
- seen
 - document, 35
- selection
 - feature, 50
 - term, 50
- selective
 - expansion, 71
- semantic
 - frame, 108
 - matching, 69
- Semi-Thue System, 74
- sensitive
 - context, 74
- sentence, 19, 74
 - ambiguous, 78
- sentential
 - form, 74
- set
 - difference, 90
 - synonym, 16
 - train, 34
 - union, 90
- shallow
 - linguistics, 64
- side
 - left-hand, 73, 94
 - right-hand, 73, 94
- significant
 - term, 70
- similarity
 - function, 17
 - NP, 24
 - VP, 30
 - word, 17
- simple Bayesian
 - classification, 39
- simultaneous
 - grammar, 105
- single
 - keyword, 7
 - single-key
 - search, 7
 - small
 - classes, 35
 - sparse
 - feature set, 36
 - split
 - nondeterministic, 88
 - sqrt
 - weighting, 42
 - standard
 - translation, 106
 - stem, 13
 - stemmer
 - KSTEM, 13
 - Porter, 13
 - stemming, 10, 13
 - stoplist, 50
 - strength, 37
 - string
 - search, 8
 - strongly
 - equivalent, 76
 - typed, 85
 - structural
 - ambiguity, 24
 - structure
 - discourse, 19
 - subcategorization, 25, 27
 - subject, 27
 - substitution
 - consistent, 84
 - supervised
 - learning, 33
 - symbol
 - initial, 73
 - nonterminal, 73, 94
 - terminal, 73, 94
 - synonym, 15
 - near, 15
 - periphrastic, 17
 - set, 16
 - synset, 16, 70
 - syntactic
 - matching, 69
 - variation, 9, 22
 - syntax
 - abstract, 108
 - analysis, 66
 - concrete, 108
 - frame, 68
 - rule, 94
 - synthesis, 77
 - tag, 10
 - tag-set, 10
 - tagger, 24
 - tagging, 10
 - part-of-speech, 10
 - taxonomy, 15
 - technology
 - linguistic, 3
 - term, 33
 - clustering, 52
 - extraction, 33
 - frequency, 50
 - selection, 50
 - significant, 70
 - terminal
 - production, 74
 - symbol, 73, 94
 - terminology
 - linguistic, 10
 - text
 - categorization, 31
 - document, 5
 - processing, 3
 - thesaurus, 5, 7, 8
 - threshold, 17, 34, 40, 53
 - range, 42
 - token, 10
 - tokenisation, 10
 - Top-Down
 - parsing, 80
 - topic, 28
 - topicalization, 28
 - TOSCA
 - project, 89
 - train
 - set, 34
 - training
 - phase, 34
 - transducer, 66, 105
 - CF, 106
 - Fibonacci, 107

- invertible, 110
- nondeterministic, 105
- transduction
 - compositional, 66
 - part, 106
- transfer
 - function, 105
- translation, 105, 106
 - standard, 106
- tree
 - operator, 108
 - parse, 77
- treebank, 10
- type, 10
- typed
 - strongly, 85
- uncertainty, 51
- union
 - operator, 94
 - set, 90
- unix, 7
- unmester, 66
- unnesting, 26, 29, 66, 68
- unseen
 - document, 35
- variable, 94
- variation
 - lexical, 8
 - linguistic, 8
 - morphological, 8
 - syntactic, 9, 22
- vector
 - weight, 38
- verb, 19
 - phrase, 27
- verbal
 - clause, 27
- vocabulary
 - growth, 37
- VP, 27
 - frame, 30
 - similarity, 30
- VWG, 76
- W
 - grammar, 88
- weakly
 - equivalent, 76
- weight
 - vector, 38
- weighting
 - binary, 42
 - keyword, 7
 - linear, 42
 - sqrt, 42
- wh-question, 9
- window, 8
- Winnow
 - algorithm, 40
 - balanced, 41
 - heuristic, 42
 - positive, 41
- word
 - compound, 12
 - order, 28
 - similarity, 17
- word sense disambiguation, 71
- word-order
 - normalization, 29
- Work-Flow
 - Management, 31