

Index

- - [λ abstraction \(10.4.2\)](#)
 - [α-conversion \(10.4.2\)](#)
 - [α equivalents \(10.4.2\)](#)
 - [λ operator \(10.4.2\)](#)
 - [β-reduction \(10.4.2\)](#)
 - [\(phrasal\) projections \(9.3.2\)](#)
- A**
 - [accuracy \(6.3.2\)](#)
 - [address \(10.4.5\)](#)
 - [adjectives \(5.2.6\)](#)
 - [adverbs \(5.2.6\)](#)
 - [agreement \(9.1.1\)](#)
 - [alphabetic variants \(10.4.2\)](#)
 - [anaphora resolution \(1.5.2\)](#)
 - [anaphoric antecedent \(10.5.1\)](#)
 - [antecedent \(1.5.2\)](#)
 - [antonymy \(2.5.3\)](#)
 - [appending \(1.2.1\)](#)
 - [appropriate \(9.5\)](#)
 - [argument \(10.2\)](#)
 - [arity \(10.3.3\)](#)
 - [articles \(5.2.6\)](#)
 - [assignment \(10.3.5\)](#)
 - [associative array \(5.3\)](#)
 - [assumption \(10.2\)](#)
 - [atomic \(9.1.3\)](#)
 - [attribute value matrix \(9.1.3\)](#)
 - [auxiliaries \(9.3.3\)](#)
 - [auxiliary \(9.1.3\)](#)
- B**
 - [backoff \(5.4.3\)](#)
 - [backtracks \(8.4.1\)](#)
 - [base case \(4.7.1\)](#)
 - [basic types \(10.3.1\)](#)
 - [bigrams \(1.3.3\)](#)
 - [binary predicate \(10.3.1\)](#)
 - [bind \(10.3.1\)](#)
 - [binding operators \(10.4.5\)](#)
 - [binning \(6.5.3\)](#)
 - [BIO Format \(7.8\)](#)
 - [Bold \(1.9\)](#)
 - [boolean \(9.1.3\)](#)
 - [boolean operators \(10.2\)](#)
 - [bottom-up \(4.7.3\)](#)

bottom-up parsing (8.4.1)

bound (10.3.1)

bound (10.3.1)

breakpoints (4.6.4)

C

call-by-value (4.4.2)

call (1.1.4)

call structure (4.7.3)

Catalan numbers (8.6.2)

characteristic function (10.3.4)

chart (8.4.4)

chart parsing (8.4.4)

child (7.4.2)

chink (7.2.5)

chink (7.8)

chunk grammar (7.2.1)

chunk (7.2)

chunking (7.2)

class label (6.1)

Classification (6.1)

closed class (5.7.4)

closed (10.3.1)

closures (3.4.2)

code point (3.3.1)

coindex (9.2)

collocation (1.3.3)

comparative wordlist (2.4.3)

complements (8.5.1)

complete (8.9)

complex (9.1.3)

complex types (10.3.1)

components (1.5.5)

concatenation (3.2.1)

conclusion (10.2)

conditional expression (1.4.3)

conditional frequency distribution (2.2)

conditional (6.6.3)

confusion matrix (6.3.4)

consecutive classification (6.1.6)

consistent (10.1.2)

constituent (8.2.1)

constituent structure (8.2.1)

control (1.4)

control structure (1.4.3)

conversion specifiers (3.9.2)

Cooper storage (10.4.5)

coordinate structure (8.2.1)

copy (4.1.1)

coreferential (10.3.1)

corpora (2)
Corpus Linguistics (2.7)
cross-validation (6.3.5)

D

data intensive (1.1)
debugger (4.6.4)
decision nodes (6.4)
decision stump (6.4)
decision tree (6.4)
decoding (3.3.1)
defensive programming (4.4.4)
dependents (8.5)
determiners (5.2.6)
dev-test (6.1.2)
development set (6.1.2)
dialogue acts (6.2.2)
dictionary (5.3)
dictionary (5.3.2)
directed acyclic graphs (9.2)
discourse (10.5)
discourse referents (10.5.1)
discourse representation structure (10.5.1)
dispersion plot (1.1.3)
divide-and-conquer (4.7)
docstring (4.4)
doctest block (4.4.6)
domain (10.3.4)
DRS conditions (10.5.1)
duck typing (7.4.3)
dynamic programming (8.4.4)

E

empiricism (1.2)
encode (6.1.1)
encoding (3.3.1)
entails (2.5.3)
equivalent (9.2)
error analysis (6.1.2)
evaluation set (6.3.1)
existential quantifier (10.3.1)
Expected Likelihood Estimation (6.5.2)
export (3.9.2)

F

F-Measure (6.3.3)
F-Score (6.3.3)
f-structure (9.5)
False negatives (6.3.3)
False positives (6.3.3)
feature extractor (6.1.1)
feature (9.1.2)

feature path (9.2)
feature set (6.1.1)
feature structures (9.1)
features (6.1.1)
fields (4.2.2)
filler (9.3.4)
folds (6.3.5)
formal language theory (1.1)
format string (3.9.2)
free (10.3.1)
frequency distribution (1.3.1)
function (2.3.2)

G

gaps (9.3.4)
gazetteer (7.5)
generalized quantifiers (10.8)
generative grammar (8.8)
generative (6.6.3)
generator expression (4.2.3)
gerund (5.7.1)
Glue semantics (10.7)
glyphs (3.3.1)
gold standard (5.4.4)
graphs (4.8.2)
greedy sequence classification (6.1.6)

H

hapaxes (1.3.1)
hash array (5.3)
head features (9.6)
head (8.5)
headword (2.4)
Heldout Estimation (6.5.2)
Hidden Markov Models (6.1.7)
Hole semantics (10.7)
holonyms (2.5.3)
homonyms (2.4)
hyponyms (2.5.2)

I

idealism (1.2)
identifiers (1.2.3)
immediate constituents (8.2.1)
immutable (4.2.2)
inconsistent (10.1.2)
indented code block (1.1.4)
independence assumption (6.5.1)
index (1.2.2)
inference (10.2)
Information Extraction (7.1)
information gain (6.4.1)

Inline annotation (11.3.5)
interpreter (1.1.1)
IOB tags (7.2.6)
iterative optimization (6.6)

J

joint classifier (6.1.6)
joint-feature (6.6.1)

K

Kappa (11.2.2)
key (5.3.2)
keyword arguments (4.5.4)
Kleene closures (3.4.2)

L

lambda expressions (4.5.1)
latent semantic analysis (4.8.4)
leaf nodes (6.4)
left-corner (8.4.3)
left-corner parser (8.4.3)
left-recursive (8.3.3)
lemma (2.4)
letter trie (4.7.1)
lexical acquisition (8.9)
lexical categories (5)
lexical entry (2.4)
lexical relations (2.5.3)
lexicon (6.10)
LGB rule (4.4.3)
library (2.3.3)
licensed (9.3.4)
likelihood ratios (6.1.1)
Linear-Chain Conditional Random Field Models (6.1.7)
list (1.2.1)
local variables (2.3.2)
logical constants (10.3.1)
logical form (10.2)

M

machine translation (1.5.3)
mapping (5.3)
maximal projection (9.3.2)
Maximum Entropy (6.6)
Maximum Entropy Markov Models (6.1.7)
Maximum Entropy principle (6.6.2)
meronyms (2.5.3)
methods (2.3.2)
modals (5.2.6)
model checking (10.3.5)
model (10.1.2)
models (6.7)

module (2.3.3)
morpho-syntactic (5.7.5)
morphological analysis (5.7.5)
multiword expression (3.11)
mutable (4.2.2)

N

n-gram tagger (5.5.3)
naïve Bayes assumption (6.5.1)
naïve Bayes (6.5)
named entity detection (7.1.1)
named entity recognition (7.5)
Negrito (1.9)
newlines (3.1.5)
NLTK Data Repository (11.6.3)
non-logical constants (10.3.1)
non-standard words (3.6.2)
normalized (3.6)
noun phrase chunking (7.2.1)
noun phrase (1.2)
NP-chunking (7.2.1)

O

objective function (3.8.2)
open class (5.7.4)
open formula (10.3.1)
out-of-vocabulary (5.5.5)
overfit (6.4.1)
overfitting (6.1.2)

P

package (2.3.3)
parameter (1.1.4)
parameters (6.5.5)
parent (7.4.2)
parser (8.4)
part-of-speech tagging (5)
partial information (9.2.1)
parts of speech (5)
personal pronouns (5.2.6)
phonology (1.1)
phrasal level (9.3.2)
POS-tagger (5.1)
POS-tagging (5)
pre-sort (4.7)
Precision (6.3.3)
precision/recall trade-off (5.5.3)
predicates (10.3.1)
prepositional phrase attachment ambiguity (8.3.1)
prepositional phrase (8.2.1)
present participle (5.7.1)
principle of compositionality (1.1)

prior probability (6.5)
probabilistic context free grammar (8.6.3)
productions (8.1.1)
projective (8.5)
proof goal (10.3.2)
Propositional logic (10.2)
propositional symbols (10.2)
prune (6.4.1)

Q

question answering (1.5.3)

R

rationalism (1.2)
raw string (3.4.2)
realism (1.2)
Recall (6.3.3)
recognizing (8.4.4)
record (4.2.2)
recursion (4.7.1)
recursive (8.3.3)
reduce (8.4.2)
reentrancy (9.2)
refactor (4.4.5)
regression testing (4.6.5)
relation detection (7.1.1)
relational operators (1.4.1)
return value (2.3.2)
root element (11.4.3)
root node (6.4)
runtime error (1.2.2)

S

S-Retrieval (10.4.5)
satisfies (10.3.5)
scope (10.3.7)
segmentation (3.8)
semantic role labeling (1.5.2)
sequence classifier (6.1.6)
sequence (3.2.6)
shift (8.4.2)
shift-reduce parser (8.4.2)
siblings (7.4.2)
signature (10.3.1)
slash categories (9.3.4)
slicing (1.2.2)
smoothing (6.5.2)
stack trace (4.6.4)
standoff annotation (11.2.3)
standoff annotation (11.3.5)
start-symbol (8.3.1)
stopwords (2.4.1)

string formatting expressions (3.9.2)
string (3.2)
strings (1.2.4)
structurally ambiguous (8.3.1)
structure sharing (9.2)
structured data (7.1)
stylistics (2.1.3)
subcategorized (8.5.1)
subsumes (9.2.1)
subsumption (9.2.1)
subtype (9.5)
supervised (6.1)
Swadesh wordlists (2.4.3)
symbolic logic (1.1)
synonyms (2.5.1)
synset (2.5.1)
syntax error (1.1.1)

T

T9 (3.4.2)
tag (9.2)
tag patterns (7.2.2)
tagged (5.2.2)
tagging (5)
tagset (5)
terms (10.3.1)
test set (6.1.1)
test set (6.3.1)
text alignment (1.5.4)
textonyms (3.4.2)
token (1.1.4)
tokenization (3.1.1)
top-down (4.7.3)
top-down parsing (8.4.1)
total likelihood (6.6)
train (5.5.1)
training (5.5.1)
training set (6.1.1)
training set (6.1.2)
transitive verbs (8.5.1)
tree (7.4.2)
True negatives (6.3.3)
True positives (6.3.3)
truth-conditions (10.2)
tuple (4.2)
Turing Test (1.5.5)
Type I errors (6.3.3)
Type II errors (6.3.3)
type-raising (10.4.3)
Typed feature structures (9.5)

types (10.3.1)

U

unary predicate (10.3.1)
unbounded dependency construction (9.3.4)
underspecified (9.1.2)
unification (9.2.1)
unique beginners (2.5.2)
universal quantifier (10.3.1)
unseen (6.7)
unstructured data (7.1)

V

valencies (8.5.1)
valid (10.2)
validity (11.4.1)
valuation function (10.3.4)
value (5.3.2)
variable (1.2.3)
verb phrase (8.2.1)

W

weights (6.5.5)
well formed formulas (10.2)
well formed (11.4.1)
well-formed substring table (8.4.4)
wildcard (3.4.1)
word classes (5)
word sense disambiguation (1.5.1)
word type (1.1.4)
WordNet (2.5)

X

XML attribute (11.4.1)
XML element (11.4.1)

Z

zero projection (9.3.2)

About this document...

This is a chapter from *Natural Language Processing with Python*, by [Steven Bird](#), [Ewan Klein](#) and [Edward Loper](#), Copyright © 2009 the authors. It is distributed with the *Natural Language Toolkit* [<http://www.nltk.org/>], Version 2.0.1rc1, under the terms of the *Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 United States License* [<http://creativecommons.org/licenses/by-nc-nd/3.0/us/>].

This document was built on Mon 15 Oct 2012 16:46:09 EST