

Index

A

acoustic model 170
 acoustic signal 213
 active learning (AL) 82
 Adobe Font Development Kit 11
 ambiguity 142-143, 146, 152, 162, 210, 284
 Anwesan 302-303, 309-312
 Apache Tika 276
 Artificial Intelligence (AI) 329
 Automatic Speech Recognition (ASR) 169-171, 208

B

bandpass filtering 234
 Bangla 42-43, 47, 51, 117, 137, 171
 adjectives 43
 nouns 42, 58, 75
 postpositions 47
 pronouns 43
 verbs 44-45, 47, 166
 Bangladesh Open Source Network (BdOSN) 237
 Bangladesh Visually Impaired People's Society (BVIPS) 31
 Bangla Word Dictionary 35, 37, 50-51, 64, 75-77
 Bengali virtual keyboard 312
 bipartite graph matching 156, 161
 BLEU 75, 79, 81, 83-89, 91, 93, 95
 Blind Education and Rehabilitation Development Organization (BERDO) 31
 blog 316-317, 334, 344, 346-347, 349, 351-355, 357, 364, 367-368
 bootstrapping 81, 365
 Brahmi 209, 247
 Braille Embosser 19, 25, 33
 Braille system 17-19, 23, 28-33
 burst frequency 265, 268, 270

C

candrabinu 8, 10
 case folding 278
 cepstral analysis 178-179, 216
 cepstrum 178, 216
chandrabinu 210
 channel normalization 176
 character fitting 1-2, 10
 character mapping 21-22, 25, 28
 character set 1-2, 5-8, 96-97, 121
 classifier 80-82, 96, 98-100, 108-109, 113-115, 141, 170, 175, 180-183, 185-186, 189, 191, 193, 199-204, 206, 234, 248, 281-282, 332, 334, 336, 365
 clause marker 157
 coarticulation effect 170, 237
 condition-action rule 39
 conjunct 7, 31, 157, 210, 353
 consonant cluster 172, 210, 222, 248
 constituent 144
 constraint graph (CG) 156
 constraint-satisfaction 155
 Context-Free Grammar (CFG) 141, 144, 154
 Context Sensitive Grammar (CSG) 141
 Contrast in Identical Environment (CIE) 249
 cross-correlation 118, 121, 123-125, 127
 Cross-Language Information Retrieval (CLIR) 290
 curvelet coefficient 105

D

decompounding 284-285, 298, 300
 deconversion 35, 41, 50, 76-77
 Demand Group 156, 160-161
 demisyllable 217, 219, 226, 245
 dental consonant 246-247, 250, 263, 265, 268, 270
 Deterministic Finite Automata (DFA) 16, 24
 diagramming 138

Index

dialect 171
digital signal processing (DSP) 175
diphone 219, 226
diphthong 318
disambiguation 83, 154-155, 167, 335, 346, 348,
362-364
Discrete Fourier Transform (DFT) 177
Dr. Sentiment 332
DSpace 206, 309-310, 312

E

emotion recognition 347
EnConverter (EnCo) 39
Europarl 81-82, 84, 87
eXtensible Markup Language (XML) 143, 316, 319

F

Field Programmable Gate Array (FPGA) 18
filter bank 176-177
Freeman chain code 105
fringe-adjusted filter (FAF) 126
fundamental frequency 176, 178, 214, 216, 219,
222-224, 242, 248-255, 269, 271

G

General Inquirer 329, 345, 349, 367
glyph set 7
granularity 290, 316, 323

H

Heaps' law 303-304, 306
hybrid feature 208
hypernasality 220, 228, 234, 245

I

inflection 37, 42, 47-57, 59-65, 74, 76-77
morphology 58, 281, 298
inflectional language 96-97, 278
Information Retrieval (IR) 273-274
input buffer 26, 28-30, 149
inserting rule 22
intensifier 353
International Components for Unicode (ICU) 277
International Phonetic Alphabet (IPA) 171
International Standards Organisation (ISO) 330
intonation 210, 214-215, 225, 227, 237, 239-240
inverse document frequency (idf) 286

J

Java 18, 30, 297, 310
joining script 11
Joint Power Spectrum (JPS) 122
Joint Transform Correlation (JTC) 118

K

kappa coefficient 348, 354
Karaka 156, 158-162
kerning 10-11
Klatt synthesizer 229
knowledge base (KB) 36

L

Label Accuracy Score (LS) 163
Labeled Attachment Score (LAS) 163
language model adaptation 81, 94
cross-domain 81
dynamic 81
Language Technology (LT) 213
left factoring 137, 139, 143, 146
leftness rule 160
lemmatization 279, 300
Letter-To-Sound (LTS) 226
Lexical Functional Grammar (LFG) 141
lexical richness 302-309, 312-314
ligature 277
Linear Predictive Coding (LPC) 216, 223
linear regression (LR) 186-187
line-casting 6, 10
Linguistic Data Consortium 84, 87, 92
Linotron 202 7
Linotype Bengali 2-4, 6, 11-12, 14
local feature 170, 187-188, 191-192, 208
logarithmic distance 251, 268-269
Log Magnitude Approximation (LMA) 225
longest run 99, 102
Lucene 296-297, 310-311

M

machine translation (MT) 79, 290
mapping table 21, 25-26, 28
matra 98-99, 101, 106-108, 120, 284
Measure of Agreement on Set-valued Items
(MASI) 347, 352-353
mel-filter 177
Mel Frequency Cepstral Coefficient
(MFCC) 169-170, 176, 185

Morpheme Extraction Task (MET) 294
 morpheme induction 281-283, 295, 299
 morphology 43, 46-47, 58-59, 63-64, 78, 89, 107-108, 139, 166, 280-281, 298-299
 MPQA corpus 348
 Multilayer Neural Network (MLN) 182, 185
 Murty Classical Library of India 4

N

nasalization 210, 212-213, 215, 220-221, 227-239, 241-245
 nasal-oral distinction 211
 natural language 39-40, 50, 79, 83, 94-95, 137-138, 140, 143, 152-155, 166, 207, 241, 274, 290-291, 293, 299-300, 302, 304, 314-317, 325-329, 338, 342-347, 350, 358, 365-367
 Natural Language Processing (NLP) 79, 143, 274, 316-317, 329, 346
n-gram 87, 92, 94-95, 278, 295, 300
 numerals 7, 97, 103-104, 114-115

O

OpenType 5
 opinion mining 315-316, 323, 326, 328-330, 343-344, 348, 365
 Optical Character Recognition (OCR) 14-15, 97-99, 108, 114-117, 120, 134-135, 277, 294
 Out-Of-Vocabulary (OOV) 170, 180

P

Page-Rank algorithm 340
 Paninian grammar 156, 166
 parameter transform 175
 parser 39, 137-143, 146-147, 149-157, 163, 165-167, 276
 parse table 137, 148
 parse tree 138-140, 142-143, 147-148, 150-151, 154
 parsing 138
 bottom-up 140-142, 152, 154
 dependency 155, 166-167
 recursive descent 140
 shift-reduce 140
 top-down 137, 142, 154
 parts of speech (POS) 138, 316
 phoneme 171-174, 193, 210-211, 214, 217, 219-220, 222-224, 241-242, 245, 247-249, 265, 268-269, 272
 phoneme accuracy 180

Phoneme Correct Rate (PCR) 180
 phoneme recognition 169, 181-183, 185-187, 206-207, 244, 249, 270
 phonetic keyboard 7, 14
 Pica Bengali 6
 Point-wise Mutual Information (PMI) 349
 polarity 325, 327-328, 330-333, 335-337, 343-345
 polysemy 275-276
 pooling 291
 power spectrum 118, 122, 127, 222-223, 241, 244, 248, 265, 271-272
 Predicate Preserving Parser (PPP) 141
 prosody 210, 214, 217, 220, 224-225, 227, 237
 Pseudo-Relevance Feedback (PRF) 275

Q

Quadratic Center of Mass (QCOM) 107, 111
 quad-tree 103
 Query Expansion (QE) 288

R

Rachanabali, Rabindra 302-304, 306-307, 309-310, 312
 real-time 118, 125, 132, 134, 237-238
 reduplication 76, 352
 Relevance Feedback (RF) 275, 288
 reph 10
 replacing rule 21
 retroflex 250, 265-268, 270
 reverse translation 85-86
 Rhetorical Structure Theory (RST) 352
 Robertson Selection Value (RSV) 289
 rule engine 24-26, 30

S

Sanskrit 156, 172, 209, 247
 Sanyal, Narayan 328, 341
 segmentation 99
 selective sampling 82, 89
 self-training 80, 82, 84
 semantic analysis 39-40, 48, 50, 60, 67-72, 152
 semi-supervised learning 79-81, 87-89, 92-93, 95
 sentiment analysis 315-317, 323, 325-331, 333, 335, 337-338, 341-345, 350, 366-367
 Sentiment Tag Extraction Program (STEP) 350
 source language 80-82, 84, 89, 361
 spectral analysis 175-177, 251, 265, 272

Index

spectral shaping 175-176
speech analysis 209, 213, 215-216, 221-223, 236-238, 245, 248-249
speech perception 215, 220, 245, 249
speech production 209, 213-214, 223, 241, 268, 270
speech signal 170, 175-178, 180-181, 188, 193, 199-200, 202, 208, 213, 215-216, 219, 221-222, 224, 234, 244, 272
speech synthesis 171, 207, 212-213, 215, 217, 222-227, 238-240, 242, 244-245
 concatenative 217, 219, 224, 226, 240, 242
 formant 217, 219
Statistical Machine Translation (SMT) 79-80, 95
stemming 279-280, 294-295, 301
stopword 278-279, 291, 294, 296-297
stroke 100
stroke weight 4, 8
structurization 331, 337-338
synset 331, 348, 358-363
syntactic analysis 50, 67

T

Tagore, Rabindranath 302, 304, 308-309, 312
tagset 331
target span 346, 352, 357, 368
Tense, Aspect, and Modality (TAM) 158
term conflation 279, 299
TextBraille 19
Text-To-Speech (TTS) 220
thinning 108
tokenization 278, 297, 300
transductive learning 82, 95, 317
transliteration 5, 7, 19, 33, 290, 296, 312
triphone 170, 172, 174-175, 189, 191, 193, 199

typeface 1-5, 7-8, 12, 14
 digital 1-3, 7, 12
 dimensions 7

U

Unicode 2, 7, 13-15, 17, 19, 24, 26, 31, 33, 226, 236-237, 277, 309
Unified Modeling Language (UML) 29-30
Universal Networking Language (UNL) 35-36, 78
universal word (UW) 36, 38, 50

V

VanderLugt filter 118
Vector Space Model (VSM) 286, 337
verb root 51-52, 56, 60, 65, 69, 72
Vertical Aspect Ratio (VAR) 107
vocal fold 215, 247, 251
vocal tract 178, 213-216, 219, 241, 245, 247
VOLT 11, 15
vowel duration 232, 245, 250, 268
vowel independence 230

W

Wavelet Transform (WT) 216, 223
Wilkins, Charles (1778) 6, 210
word alignment 81
WordNet 317-318, 325, 327, 331, 346-348, 350, 357-364, 366-368

Z

zero-order 122, 125, 127-128