

Contents

Part I Concepts and Technology

1	Natural Language Processing	3
1.1	Introduction	3
1.2	Human Language and Intelligence	4
1.3	Linguistic Levels of Human Language	6
1.4	Human Language Ambiguity	7
1.5	A Brief History of NLP	8
1.5.1	First Stage: Machine Translation (Before the 1960s)	8
1.5.2	Second Stage: Early AI on NLP (1960s–1970s)	8
1.5.3	Third Stage: Grammatical Logic on NLP (1970s–1980s)	9
1.5.4	Fourth Stage: AI and Machine Learning (1980s–2000s)	9
1.5.5	Fifth Stage: Rise of BERT, Transformer, ChatGPT, and LLMs (2000s–Present)	9
1.6	NLP and AI	11
1.7	Main Components of NLP	12
1.8	Natural Language Understanding (NLU)	12
1.8.1	Speech Recognition	13
1.8.2	Syntax Analysis	14
1.8.3	Semantic Analysis	14
1.8.4	Pragmatic Analysis	14
1.9	Potential Applications of NLP	14
1.9.1	Machine Translation (MT)	14
1.9.2	Information Extraction (IE)	15
1.9.3	Information Retrieval (IR)	15
1.9.4	Sentiment Analysis	16
1.9.5	Question-Answering (Q&A) Chatbots	16
	References	17

2	N-Gram Language Model	19
2.1	Introduction	19
2.2	N-Gram Language Model	21
2.2.1	Basic NLP Terminology	23
2.2.2	Language Modeling and Chain Rule	24
2.3	Markov Chain in N-Gram Model	26
2.4	Example: The Adventures of Sherlock Holmes	27
2.5	Shannon's Method in N-Gram Model	31
2.6	Language Model Evaluation and Smoothing Techniques	34
2.6.1	Perplexity	34
2.6.2	Extrinsic Evaluation Scheme	35
2.6.3	Zero Counts Problems	36
2.6.4	Smoothing Techniques	36
2.6.5	Laplace (Add-One) Smoothing	37
2.6.6	Add- k Smoothing	38
2.6.7	Backoff and Interpolation Smoothing	40
2.6.8	Good Turing Smoothing	41
	References	42
3	Part-of-Speech (POS) Tagging	45
3.1	What Is Part of Speech (POS)?	45
3.1.1	Nine Major POS in the English Language	45
3.2	POS Tagging	46
3.2.1	What Is POS Tagging in Linguistics?	46
3.2.2	What Is POS Tagging in NLP?	46
3.2.3	POS Tags Used in the PENN Treebank Project	47
3.2.4	Why Do We Care About POS in NLP?	49
3.3	Major Components in NLU	50
3.3.1	Computational Linguistics and POS	51
3.3.2	POS and Semantic Meaning	52
3.3.3	Morphological and Syntactic Definition of POS	52
3.4	Nine Key POS in English	53
3.4.1	English Word Classes	53
3.4.2	What Is a Preposition?	54
3.4.3	What Is a Conjunction?	54
3.4.4	What Is a Pronoun?	55
3.4.5	What Is a Verb?	56
3.5	Different Types of POS Tagset	58
3.5.1	What Is Tagset?	58
3.5.2	Ambiguous in POS Tags	59
3.5.3	POS Tagging Using Knowledge	59
3.6	Approaches for POS Tagging	60
3.6.1	Rule-Based Approach POS Tagging	60
3.6.2	Example of Rule-Based POS Tagging	61
3.6.3	Example of Stochastic-Based POS Tagging	62

3.6.4	Hybrid Approach for POS Tagging Using Brill's Taggers. .	63
3.6.5	What Is Transformation-Based Learning?.....	63
3.6.6	Hybrid POS Tagging: Brill's Tagger	63
3.6.7	Learning Brill's Tagger Transformations.....	64
3.7	Taggers Evaluations.....	65
3.7.1	How Good Is an POS Tagging Algorithm?	66
	References.....	67
4	Syntax and Parsing.....	69
4.1	Introduction and Motivation	69
4.2	Syntax Analysis	70
4.2.1	What Is Syntax.....	70
4.2.2	Syntactic Rules.....	70
4.2.3	Common Syntactic Patterns.....	71
4.2.4	Importance of Syntax and Parsing in NLP	72
4.3	Types of Constituents in Sentences	72
4.3.1	What Is Constituent?	72
4.3.2	Kinds of Constituents.....	74
4.3.3	Complexity on Simple Constituents	76
4.3.4	Verb Phrase Subcategorization	76
4.3.5	The Role of Lexicon in Parsing.....	77
4.3.6	Recursion in Grammar Rules.....	78
4.4	Context-Free Grammar (CFG).....	79
4.4.1	What Is Context Free Language (CFL)?	79
4.4.2	What Is Context Free Grammar (CFG)?	79
4.4.3	Major Components of CFG	80
4.4.4	Derivations Using CFG	80
4.5	CFG Parsing.....	81
4.5.1	Morphological Parsing.....	81
4.5.2	Phonological Parsing	82
4.5.3	Syntactic Parsing	82
4.5.4	Parsing as a Kind of Tree Searching	82
4.5.5	CFG for Fragment of English	83
4.5.6	Parse Tree for "Play the Piano" for Prior CFG	83
4.5.7	Top-Down Parser	83
4.5.8	Bottom-Up Parser	85
4.5.9	Control of Parsing	86
4.5.10	Pros and Cons of Top-Down vs. Bottom-Up Parsing	87
4.6	Lexical and Probabilistic Parsing.....	88
4.6.1	Why Using Probabilities in Parsing?.....	88
4.6.2	Semantics with Parsing	88
4.6.3	What Is PCFG?	89
4.6.4	A Simple Example of PCFG	90
4.6.5	Using Probabilities for Language Modelling.....	92
4.6.6	Limitations for PCFG	92

4.6.7	The Fix–Lexicalized Parsing	93
	References.....	96
5	Meaning Representation.....	99
5.1	Introduction	99
5.2	What Is Meaning?	99
5.3	Meaning Representations.....	100
5.4	Semantic Processing	101
5.5	Common Meaning Representation	102
5.5.1	First-Order Predicate Calculus.....	102
5.5.2	Semantic Networks	102
5.5.3	Conceptual Dependency Diagram	102
5.5.4	Frame-Based Representation	103
5.6	Requirements for Meaning Representation	104
5.6.1	Verifiability	104
5.6.2	Ambiguity	105
5.6.3	Vagueness.....	105
5.6.4	Canonical Forms	105
5.7	Inference	106
5.7.1	What Is Inference?.....	106
5.7.2	Example of Inferencing with FOPC	107
5.8	Fillmore’s Theory of Universal Cases	107
5.8.1	What Is Fillmore’s Theory of Universal Cases?	108
5.8.2	Major Case Roles in Fillmore’s Theory.....	109
5.8.3	Complications in Case Roles	110
5.9	First-Order Predicate Calculus.....	111
5.9.1	FOPC Representation Scheme.....	111
5.9.2	Major Elements of FOPC.....	111
5.9.3	Predicate-Argument Structure of FOPC	112
5.9.4	Meaning Representation Problems in FOPC.....	114
5.9.5	Inferencing Using FOPC	115
	References.....	117
6	Semantic Analysis	119
6.1	Introduction	119
6.1.1	What Is Semantic Analysis?.....	119
6.1.2	The Importance of Semantic Analysis in NLP	120
6.1.3	How Human Is Good in Semantic Analysis?.....	120
6.2	Lexical vs. Compositional Semantic Analysis.....	121
6.2.1	What Is Lexical Semantic Analysis?	121
6.2.2	What Is Compositional Semantic Analysis?	121
6.3	Word Senses and Relations	122
6.3.1	What Is Word Sense?.....	122
6.3.2	Types of Lexical Semantics	123
6.4	Word Sense Disambiguation	127
6.4.1	What Is Word Sense Disambiguation (WSD)?	127
6.4.2	Difficulties in WSD	127

6.4.3	Method for WSD	128
6.5	WordNet and Online Thesauri	129
6.5.1	What Is WordNet?	129
6.5.2	What Are Synsets?	130
6.5.3	Knowledge Structure of WordNet	130
6.5.4	What Are Major Lexical Relations Captured in WordNet?	131
6.5.5	Applications of WordNet and Thesauri?	132
6.6	Other Online Thesauri: MeSH	133
6.6.1	What Is MeSH?	133
6.6.2	Uses of the MeSH Ontology	134
6.7	Word Similarity and Thesaurus Methods	134
6.7.1	Introduction	134
6.7.2	Path-Based Similarity	135
6.7.3	Problems with Path-Based Similarity	136
6.7.4	Information Content Similarity	137
6.7.5	The Resnik Method	138
6.7.6	The Dekang Lin Method	138
6.7.7	The (Extended) Lesk Algorithm	139
6.8	Distributed Similarity	140
6.8.1	Distributional Models of Meaning	140
6.8.2	Word Vectors	140
6.8.3	Term-Document Matrix	140
6.8.4	Pointwise Mutual Information	142
6.8.5	Example of Computing PPMI on a Term-Context Matrix	142
6.8.6	Weighing PMI Techniques	144
6.8.7	Add-K Smoothing in PMI Computation	144
6.8.8	Context and Word Similarity Measurement	146
6.8.9	Evaluating Similarity	148
	References	149
7	Pragmatic Analysis and Discourse	151
7.1	Introduction	151
7.2	Discourse Phenomena	151
7.2.1	Coreference Resolution	152
7.2.2	Why Is It Important?	152
7.2.3	Coherence and Coreference	153
7.2.4	Importance of Coreference Relations	154
7.2.5	Entity-Based Coherence	155
7.3	Discourse Segmentation	156
7.3.1	What Is Discourse Segmentation?	156
7.3.2	Unsupervised Discourse Segmentation	156
7.3.3	Hearst's TextTiling Method	157
7.3.4	TextTiling Algorithm	158
7.3.5	Supervised Discourse Segmentation	159

7.4	Discourse Coherence	160
7.4.1	What Makes a Text Coherent?	160
7.4.2	What Is Coherence Relation?	160
7.4.3	Types of Coherence Relations	161
7.4.4	Hierarchical Structure of Discourse Coherence.	161
7.4.5	Types of REs	162
7.4.6	Features for Filtering Potential Referents	163
7.4.7	Preferences in Pronoun Interpretation	163
7.5	Algorithms for Coreference Resolution.	164
7.5.1	Introduction	164
7.5.2	Hobbs' Algorithm	164
7.5.3	Centering Algorithm	167
7.5.4	Machine-Learning Method	170
7.6	Evaluation	171
	References.	173
8	Transfer Learning and Transformer Technology.	175
8.1	What Is Transfer Learning?	175
8.2	Motivation of TL	175
8.2.1	Categories of TL	176
8.3	Solutions of TL	177
8.3.1	Instance-Based Method	178
8.3.2	Feature-Based Method.	178
8.3.3	Parameter-Based Method.	178
8.3.4	Relational-Based Method.	179
8.4	Recurrent Neural Network (RNN).	180
8.4.1	What Is RNN?	180
8.4.2	Motivation of the RNN	180
8.4.3	RNN Architecture	181
8.4.4	Long Short-Term Memory (LSTM) Network	182
8.4.5	Gate Recurrent Unit (GRU).	185
8.4.6	Bidirectional Recurrent Neural Networks (BRNNs).	187
8.5	Transformer Technology	188
8.5.1	What Is Transformer?	188
8.5.2	Transformer Architecture.	188
8.5.3	Deep into Encoder	190
8.6	BERT	192
8.6.1	What Is BERT?	192
8.6.2	Architecture of BERT	192
8.6.3	Training of BERT.	193
8.7	Other Related Transformer Technology.	194
8.7.1	Transformer-XL.	194
8.7.2	ALBERT	195
	References.	196

9	Major NLP Applications	199
9.1	Introduction	199
9.2	Information Retrieval Systems	199
9.2.1	Introduction to IR Systems	199
9.2.2	Vector Space Model in IR	200
9.2.3	Term Distribution Models in IR	202
9.2.4	Latent Semantic Indexing in IR	207
9.2.5	Discourse Segmentation in IR	209
9.3	Text Summarization Systems	213
9.3.1	Introduction to Text Summarization Systems	213
9.3.2	Text Summarization Datasets	215
9.3.3	Types of Summarization Systems	215
9.3.4	Query-focused vs. Generic Summarization Systems	216
9.3.5	Single and Multiple Document Summarization	217
9.3.6	Contemporary Text Summarization Systems	219
9.4	Question-and-Answering Systems	224
9.4.1	QA System and AI	224
9.4.2	Overview of Industrial QA Systems	229
	References	236
10	Large Language Models (LLMs) and Generative Artificial Intelligence (GenAI)	241
10.1	Introduction to LLM and GenAI	241
10.1.1	What Is a Large Language Model (LLM)?	241
10.1.2	Understanding Generative Artificial Intelligence (GenAI)	242
10.1.3	The Intersection of LLM and GenAI	243
10.1.4	The Importance of LLMs in Modern AI	244
10.2	Foundations of LLMs	245
10.2.1	Neural Network Architectures	245
10.2.2	Attention Mechanisms	245
10.2.3	The Transformer Architecture	246
10.2.4	Scaling Up: From BERT to GPT	248
10.3	Key Players in the LLM Landscape	249
10.3.1	ChatGPT by OpenAI (Current Version: GPT-4)	249
10.3.2	Pathways Language Model (PaLM) by Google DeepMind (Current Version: PaLM 2)	250
10.3.3	Large Language Model Meta AI (LLaMA) by Meta (Current Version: LLaMA 2)	251
10.3.4	Claude by Anthropic (Current Version: Claude 2)	252
10.3.5	ERNIE 3.0 Titan by Baidu	253
10.4	Applications of LLMs in GenAI	254
10.4.1	Creative Writing and Content Generation	255
10.4.2	Language Translation	256
10.4.3	Conversational AI and Chatbots	257
10.4.4	Text Summarization and Content Curation	258

10.5	Ethical Considerations and Challenges	260
10.5.1	Detecting and Mitigating Bias	260
10.5.2	Privacy and Data Security	261
10.5.3	The Spread of Misinformation.....	262
10.5.4	Ethical Guidelines for LLM Deployment	263
10.6	Future Outlook and Research.....	264
10.6.1	Current Trends in LLMs and GenAI	264
10.6.2	The Future of Creativity in AI	265
10.6.3	The Role of LLMs in AI Ethics	267
10.6.4	The Path Forward: Research and Development.....	268
	References.....	271

Part II Natural Language Processing Workshops with Python Implementation in 14 Hours

11	Workshop#1: Basics of Natural Language Toolkit (Hour 1–2).	277
11.1	Introduction	277
11.2	What Is Natural Language Toolkit (NLTK)?.....	278
11.3	A Simple Text Tokenization Example Using NLTK	279
11.4	How to Install NLTK?	280
11.5	Why Using Python for NLP?.....	281
11.6	NLTK with Basic Text Processing in NLP	283
11.7	Simple Text Analysis with NLTK	284
11.8	Text Analysis Using Lexical Dispersion Plot	288
11.8.1	What Is a Lexical Dispersion Plot?	288
11.8.2	Lexical Dispersion Plot over Context Using Sense and Sensibility.....	289
11.8.3	Lexical Dispersion Plot Over Time Using Inaugural Address Corpus	290
11.9	Tokenization in NLP with NLTK.....	290
11.9.1	What Is Tokenization in NLP?.....	290
11.9.2	Different Between Tokenize() vs Split()	291
11.9.3	Count Distinct Tokens	292
11.9.4	Lexical Diversity	294
11.10	Basic Statistical Tools in NLTK.....	296
11.10.1	Frequency Distribution—FreqDist()	296
11.11	Do the Same Plot with <i>Cumulative</i> Mode	298
11.11.1	Rare Words—Hapax.....	299
11.11.2	Collocations	300
	References.....	302
12	Workshop#2: N-Grams Modeling with Natural Language Toolkit (Hour 3–4).	303
12.1	Introduction	303
12.2	What Is N-Gram?.....	304
12.3	Applications of N-Grams in NLP	304

12.4	Generation of N-Grams in NLTK	304
12.5	Generation of N-Grams Statistics	307
12.6	spaCy in NLP.....	313
12.6.1	What Is spaCy?	313
12.7	How to Install spaCy?	314
12.8	Tokenization Using spaCy	314
	References.....	320
13	Workshop#3: Part-of-Speech Tagging Using Natural Language Toolkit (Hour 5–6).....	323
13.1	Introduction	323
13.2	A Revisit on Tokenization with NLTK	324
13.3	Stemming Using NLTK	326
13.3.1	What Is Stemming?	327
13.3.2	Why Stemming?.....	327
13.3.3	How to Perform Stemming?.....	327
13.4	Stop-Words Removal with NLTK	330
13.4.1	What Are Stop-Words?.....	330
13.4.2	NLTK Stop-Words List	331
13.4.3	Try Some Texts.....	332
13.4.4	Create your Own Stop-Words	333
13.5	Text Analysis with NLTK	334
13.6	Integration with WordCloud	336
13.6.1	What Is WordCloud?	336
13.7	POS Tagging with NLTK.....	338
13.7.1	What Is POS Tagging?.....	338
13.7.2	Universal POS Tagset.....	339
13.7.3	PENN Treebank Tagset (English & Chinese)	339
13.7.4	Applications of POS Tagging.....	341
13.8	Create Own POS Tagger with NLTK.....	344
	References.....	350
14	Workshop#4 Semantic Analysis and Word Vectors Using spaCy (Hour 7–8).....	353
14.1	Introduction	353
14.2	What Are Word Vectors?	353
14.3	Understanding Word Vectors	354
14.3.1	Example: A Simple Word Vector	354
14.4	A Taste of Word Vectors.....	356
14.5	Analogies and Vector Operations.....	359
14.6	How to Create Word Vectors?	360
14.7	spaCy Pretrained Word Vectors	360
14.8	Similarity Method in Semantic Analysis	363
14.9	Advanced Semantic Similarity Methods with spaCy	366
14.9.1	Understanding Semantic Similarity	366
14.9.2	Euclidean Distance.....	366

14.9.3	Cosine Distance and Cosine Similarity	368
14.9.4	Categorizing Text with Semantic Similarity	370
14.9.5	Extracting Key Phrases	371
14.9.6	Extracting and Comparing Named Entities	372
	References.	374
15	Workshop#5: Sentiment Analysis and Text Classification (Hour 9–10)	375
15.1	Introduction	375
15.2	Text Classification with spaCy and LSTM Technology	375
15.3	Technical Requirements.	376
15.4	Text Classification in a Nutshell	376
15.4.1	What Is Text Classification?	376
15.4.2	Text Classification as AI Applications	377
15.5	Text Classifier with spaCy NLP Pipeline.	378
15.5.1	TextCategorizer Class	378
15.5.2	Formatting Training Data for the TextCategorizer.	380
15.5.3	System Training	383
15.5.4	System Testing	385
15.5.5	Training TextCategorizer for Multi-Label Classification.	386
15.6	Sentiment Analysis with spaCy	389
15.6.1	IMDB Large Movie Review Dataset	389
15.6.2	Explore the Dataset	390
15.6.3	Training the TextClassifier	393
15.7	Artificial Neural Network in a Nutshell.	396
15.8	An Overview of TensorFlow and Keras.	396
15.9	Sequential Modeling with LSTM Technology.	397
15.10	Keras Tokenizer in NLP.	398
15.10.1	Embedding Words	401
15.11	Movie Sentiment Analysis with LTSM Using Keras and spaCy	402
15.11.1	Step 5: Implement the LSTM Layer	406
15.11.2	Step 6: Implement the Output Layer	407
15.11.3	Step 7: System Compilation.	407
	References.	409
16	Workshop#6 Transformers with spaCy and TensorFlow (Hour 11–12).	411
16.1	Introduction	411
16.2	Technical Requirements.	412
16.3	Transformers and Transfer Learning in a Nutshell	412
16.4	Why Transformers?	413
16.5	An Overview of BERT Technology.	416
16.5.1	What Is BERT?	416
16.5.2	BERT Architecture.	417
16.5.3	BERT Input Format	418
16.5.4	How to Train BERT?	420

16.6	Transformers with TensorFlow	421
16.6.1	HuggingFace Transformers	422
16.6.2	Using the BERT Tokenizer.....	423
16.6.3	Word Vectors in BERT.....	427
16.7	Revisit Text Classification Using BERT	428
16.7.1	Data Preparation.....	428
16.7.2	Start the BERT Model Construction	429
16.8	Transformer Pipeline Technology	432
16.8.1	Transformer Pipeline for Sentiment Analysis	433
16.8.2	Transformer Pipeline for QA System.....	433
16.9	Transformer and spaCy	434
	References.....	438
17	Workshop#7 Building Chatbot with TensorFlow and Transformer Technology (Hour 13–14).....	441
17.1	Introduction	441
17.2	Technical Requirements.....	441
17.3	AI Chatbot in a Nutshell	442
17.3.1	What Is a Chatbot?.....	442
17.3.2	What Is a Wake Word in Chatbot?.....	443
17.3.3	NLP Components in a Chatbot	444
17.4	Building Movie Chatbot by Using TensorFlow and Transformer Technology.....	445
17.4.1	The Chatbot Dataset.....	445
17.4.2	Movie Dialogue Preprocessing	445
17.4.3	Tokenization of Movie Conversation.....	447
17.4.4	Filtering and Padding Process	448
17.4.5	Creation of TensorFlow Movie Dataset Object (mDS) ..	449
17.4.6	Calculate Attention Learning Weights	450
17.4.7	Multi-Head-Attention (MHAttention)	451
17.4.8	System Implementation	452
17.5	Related Works	470
	References.....	470
	Index.....	473