

Contents

1. INTRODUCTION	1
1. Introduction to Natural Language Processing	2
2. Need for Natural Language Processing (NLP)	2
3. Goals of Natural Language Processing:	3
4. Brief overview of NLP	4
5. History of NLP	5
6. Generic NLP system	7
7. Levels of NLP	8
8. Knowledge in Language processing	9
9. Ambiguity in NLP	10
10. Stages of NLP	15
11. Applications of NLP	21
Expected Questions	22
2. WORD LEVEL ANALYSIS	23
1. Morphology analysis	24
→ Survey of English Morphology	25
→ Inflectional morphology & Derivational morphology	27

2. Stemming and Lemmatization	33
→ Stemming Algorithms Examples	35
3. Regular expression	37
4. Finite Automata	41
5. Finite-State Morphological Parsing	47
6. Building a Finite-State Lexicon	48
→ Finite-State Transducers	52
→ Morphological Parsing with Finite-State Transducers	54
→ Orthographic Rules and Finite-State Transducers	57
7. Lexicon free FST Porter stemmer	60
→ Porter Stemmer	61
8. N-Grams	67
→ Language model	69
→ N-gram language model	72
→ N-gram for spelling correction	76
Expected Questions:	78
3. SYNTAX ANALYSIS	79
Syntax Analysis	80
1. Part-Of-Speech tagging(POS)	81
→ Tag set for English	82
→ Penn Treebank	85
2. Rule based POS tagging, Stochastic POS tagging	87
→ Rule-Based Part-of-Speech Tagging	88
→ Properties of Rule-Based POS Tagging	92
→ Stochastic Part of Speech Taggers	92
→ Transformation-Based Tagging	95
3. Issues	98
→ Multiple tags and multiple words	98
→ Unknown words	98
4. Introduction to CFG	99

5. Subcategorization	103
6. Movement	104
→ Parsing With Context-free Grammar	104
7. Sequence labeling	109
→ Hidden Markov Model (HMM)	111
→ HMM for POS tagging	113
→ Maximum Entropy	122
→ Conditional Random Field (CRF)	123
Expected Questions	125
4. SEMANTIC ANALYSIS	127
1. Lexical semantics	130
→ Compositional semantics	131
→ What is language understanding	131
→ Semantic analysis vs. other areas of natural language processing	132
→ Approaches to semantic analysis	132
→ Applications of semantic analysis	133
→ Why is semantic analysis difficult?	134
→ Why is semantic analysis important?	134
- 2. Attachment for Fragment of English	134
→ Phrase level Constructions	135
3. Relations among lexemes & their senses –Homonymy, Polysemy, Synonymy, Hyponymy	141
4. WordNet	142
5. Robust Word Sense Disambiguation (WSD) - Dictionary based approach.	144
Expected Questions	150
5. PRAGMATICS	151
1. Pragmatic analysis	152
Five aspects of pragmatics	155
→ Deixis:	155
→ Implicature	156

→ Presupposition	157
→ Speech Acts	158
→ Conversational Structure	159
Application that demands pragmatic understanding :	159
2. Discourse - reference resolution	160
→ Reference Resolution	161
3. Reference Phenomenon	163
4. Syntactic and Semantic Constraints on Coreference	166
Coreference.	167
Coreference distinctions	168
Coreference resolution	169
Approach to coreference resolution	170
Why Coreference Resolution is Hard	173
Coreference vs. Anaphora.	174
Application of Coreference Resolution	175
Expected Questions	176
6. APPLICATIONS (PREFERABLY FOR INDIAN REGIONAL LANGUAGES) ..	177
1. Machine Translation	178
→ Machine Translation System Approaches	178
→ Rule Based Machine Translation System	180
2. Information Retrieval	184
3. Question Answering System	187
4. Sentiment Analysis	189
→ Sentiment Classification Techniques	190
5. Text Categorization or Classification	198
→ Text Classification Techniques	198
6. Text Summarization.	204
7. Named Entity Recognition.	208
Expected Questions	212

LAB MANUAL NLP : COMPUTATIONAL LAB II	213
Experiment No 1:	214
→ Morphology is of two types:	215
→ Morphological Features:	215
Experiment No 2:	218
Experiment No 3:	219
Experiment No 4:	221
Experiment No 5	225
Experiment No 6	226
Experiment No 7	231
Experiment No 8	247
REFERENCES	250
INDEX	254