

Subject: NATURAL LANGUAGE PROCESSING

Topics to be covered, unit-wise for university examination as per the syllabus.

UNIT-1: OVERVIEW & LANG MODELLING (Chapter 1,2 – Tanveer Siddique)

1. Grammar-based (GB Theory, X-bar theory, LFG)
2. Statistical language model (n-gram, Add-one smoothing)

UNIT-2:: WORD LEVEL & SYNTACTIC ANALYSIS

3. Regular expressions
4. Morphological parsing using FST
5. POS Tagging (Rule-based, Stochastic and Hybrid taggers)
6. Constituency (Phrase-level and Sentence level constructions)
7. Parsing (top-down & bottom-up)
8. Earley parser
9. CYK parser
10. Probabilistic CFG – resolving ambiguities using PCFG

UNIT-3:: SEMANTIC ANALYSIS & DISCOURSE PROCESSING

11. Syntax-driven semantic analysis using lambda calculus
12. Thematic roles, Selectional restrictions
13. Word senses, Word Sense Disambiguation (WSD)
 - 13.1 Lesk & Resnik algorithm
 - 13.2 Naive Bayes, kNN-based
 - 13.3 Yarowsky (Bootstrapping method)
14. Cohesion
15. Reference resolution (Anaphora)
 - 15.1 Constraints on Reference
 - 15.2 Lappin Leass Algorithm
 - 15.3 Centering algorithm
 - 15.4 Mitkov's algorithm
16. Coherence relation
17. Discourse structure

UNIT-4:: NLG AND MT (Chapter 7 & 8 – Tanveer Siddique)

18. NLG architecture
19. Discourse planning, Surface realization
20. Problems in MT, Indian languages issues
21. Direct MT
22. Rule-based MT
 - 22.1 Transfer-based MT
 - 22.2 Interlingua MT
23. Corpus-based MT
 - 23.1 Statistical MT
 - 23.2 Example-based MT

UNIT-5:: IR & LEXICAL RESOURCES (Chapter 9 & 12 – Tanveer Siddique)

24. Information Retrieval Design

25. Classical IR model

- 25.1 boolean model
- 25.2 term weight model
- 25.3 probabilistic model
- 25.4 vector space model

26. Alternative model

- 26.1 Cluster model
- 26.2 Fuzzy model

27. Lexical resources

- 27.1 WordNet (<http://wordnet.princeton.edu/>)
- 27.2 FrameNet (<https://framenet.icsi.berkeley.edu/fndrupal/>)
- 27.3 POS Taggers

For Unit 2 & 3, Chapters from Jurafsky & Martin can be used to have more clarification on each topic and as additional resources.

Speech and Language Processing, Jurafsky & Martin, Pearson Education LPE, 2002

For Unit-2: Chapters from Jurafsky & Martin

Chapters 2, 3, 8, 9, 10, 12

For Unit--3: Chapters from Jurafsky & Martin

Chapters 14, 15, 16, 17, 16

If you have any another edition of above book, the chapter numbers may vary.

BEST OF LUCK!!!
