

Course No.	Type	Subject	L	T	P	Credits	CA	MS	ES	CA	ES	Pre-requisites
COCSE27	ED	Natural Language Processing	3	1	0	4						CACSC11 AI

COURSE OUTCOMES

1. Realize the fundamentals of natural language processing.
2. Recognize the basics of Language modelling and Word level analysis.
3. Understand the use of CFG and PCFG in NLP.
4. Recognize the role of semantics of sentences and pragmatic.
5. Identify the recent trends and applications of NLP.

COURSE CONTENTS:

UNIT I

Introduction: Various stages of NLP, Why NLP is hard, Why NLP is useful, Classical problems, Introduction to basic language processing – tokens, sentences, paragraphs, Spelling Correction, Morphological analysis and generation using Finite State Automata and Finite State transducer.

revisit FST and FSA

UNIT II

Language Modeling (LM): Grammar-based LM, Statistical LM, Linguistics Fundamentals, Classical approaches to NLP with knowledge bases and linguistic rules; Data Driven and Machine Learning Approaches to NLP, Text classification evaluation, relation extraction

Word Level Analysis: Unsmoothed N-grams, Evaluating N-grams, Smoothing, Interpolation and Backoff – Word Classes, Part-of-Speech Tagging, Rule-based, Stochastic and Transformation-based tagging, Issues in PoS tagging – Hidden Markov and Maximum Entropy models.

maximum entropy model revisit.

UNIT III

Syntactic Analysis: Context Free Grammars, Grammar rules for English, Treebanks, Normal Forms for grammar – Dependency Grammar – Syntactic Parsing, Ambiguity, Dynamic Programming parsing – Shallow parsing – Probabilistic CFG, Probabilistic CYK, Probabilistic Lexicalized CFGs – Feature structures, Unification of feature structures.

UNIT IV

Semantics and Pragmatics: Requirements for representation, First-Order Logic, Description Logics – Syntax-Driven Semantic analysis, Semantic attachments – Word Senses, Relations

12, 13, 14

15
wordNet

19 19.7 ND TGL
between Senses, Thematic Roles, selection restrictions – Word Sense Disambiguation, WSD using Supervised, Dictionary & Thesaurus, Bootstrapping methods – Word Similarity using Thesaurus and Distributional methods

UNIT V

Applications of NLP: Speech Processing, Speech Analysis and Modelling, Machine Translation, Information Retrieval, Text Summarization, Text Classification, Sentiment analysis and opinion mining.

SUGGESTED READINGS

1. Daniel Jurafsky, James H. Martin—Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech, Pearson Publication, 2014.
2. Christopher D. Manning and Hinrich Schütze, “Foundations of Natural Language Processing”, 6th Edition, The MIT Press Cambridge, Massachusetts London, England, 2003.
3. Steven Bird, Ewan Klein and Edward Loper, —Natural Language Processing with Python, First Edition, O'Reilly Media, 2009.
4. Hobson Lane, Cole Howard, Hannes Hapke, “Natural language processing in action” MANNING Publications, 2019.

UNIT-3

Syntactic Analysis: Context Free Grammars (12.1, 12.2), Grammar rules for English (12.3), Treebanks, (12.4) Normal Forms for grammar (12.5)

Dependency Grammar – Syntactic Parsing (NPTEL Lec 27), Ambiguity (13.1), Dynamic Programming parsing – Shallow parsing (13.5) – Probabilistic CFG (NPTEL Lec 22, 23), Probabilistic CYK (NPTEL Lec 24, 25 upto 7 min), Probabilistic Lexicalized CFGs (PDF PROVIDED, first four topics) – Feature structures, Unification of feature structures. (PDF PROVIDED, upto 13.4)

UNIT-4

Requirements for representation (15.1, 15.2), First-Order Logic (15.3), Description Logics (15.5) – Syntax-Driven Semantic analysis(?), Semantic attachments – Word Senses (18.1), Relations between Senses (18.2), WordNet (18.3) Thematic Roles, selection restrictions (19.1, 19.7) – Word Sense Disambiguation, WSD using Supervised, Dictionary & Thesaurus, Bootstrapping methods – Word Similarity using Thesaurus and Distributional methods (NPTEL Lecture 37-39)