

Program Name	B. Tech. (Computer Engineering)	Semester – VII
Course Code	R4CO4204T	
Course Title	Natural Language Processing	
Prerequisite	Machine Learning	

COURSE OUTCOMES: Students will be able to

1.	Understand the applications and analysis of NLP.
2.	Demonstrate accomplishments of knowledge and comprehension of NLP.
3.	Compare and contrast approaches to NLP.
4.	Discuss the limitations and promise of NLP.

COURSE CONTENTS

	Hrs	CO
1. Introduction to NLP: History of NLP, Generic NLP system, levels of NLP, Knowledge in language processing, Ambiguity in Natural Language, Stages in NLP, Challenges of NLP. Various Applications of NLP- Machine translation, Named Entity Recognition, question answering system, Information retrieval, Text categorization, text summarization, Sentiment Analysis and so on.	3	1
2. Word Level Analysis: Morphology in natural languages, Morphology analysis, Inflectional morphology & Derivational morphology, Regular expression, finite automata, finite state transducers (FST), Morphological parsing with FST, Lexicon free FST - Porter stemmer.	4	2
3. N-gram Language Models: The role of language models, Simple N-gram models. Estimating parameters and smoothing - Laplace Smoothing, Add-k smoothing, Kneser-Ney Smoothing, Web and Stupid Back-off, Perplexity and Entropy, Evaluating language models.	4	2
4. Part of Speech Tagging: Part of Speech Tagging, Tagset for English, Penn TreeBank Tagset for English, Rule-based Part-of-speech Tagging, Stochastic Part-of-speech Tagging, Transformation-Based Tagging, Issues –Multiple tags & words, Unknown words, class-based n-grams, Hidden Markov Models (Forward and Viterbi algorithms and EM training) for sequence labeling task of POS tagging.	4	2
5. Syntactic Analysis: Context Free grammar Constituency, Context free rules & trees, Sentence level construction - Noun Phrase, Verb phrase etc. Need of Parsing, Dependency Parsing, Parsing in case of Ambiguity, Parsing, Different Parsing Algorithms - CFG, V Probabilistic Parsing.	5	2
6. Semantics Analysis: Meaning representation, lexical semantics - sense relations, semantic roles, and primitive decomposition, WordNet - Relations among lexemes & their senses – Homonymy, Polysemy, Synonymy, Hyponymy. Ambiguity in Word Senses, Word Sense Disambiguation- Selectional restriction, machine learning approaches, dictionary-based approaches.	4	3

7.	Distributional Semantics: Distributional hypothesis, vector space models, etc. Distributed Representations: Neural Networks (NN), Backpropagation, Softmax, Hierarchical Softmax. Word Vectors: Feedforward NN, Word2Vec, GloVE, Contextualization (ELMo etc.) etc	4	3
8.	Discourse Analysis: Pragmatics and Discourse, Ambiguity in Discourse Analysis, Reference resolution, constraints on co-reference, algorithm for pronoun resolution, text coherence, discourse structure in brief.	3	4
9.	Text Classification: Preparing Data: Acquiring Text, Text Cleaning, Text Preprocessing with POS tagger, Parsers, Chunkers etc. Statistical analysis of text or sentences. Supervised approach using Classifiers such as Naive Bayes, SVM, Random Forest etc. NN-based Approaches such as LSTM-Based approach using word vectors etc. Dictionary-based or Unsupervised Approaches: Extracting Information using Corpus, evaluating statistical data, Linguistic Properties, WordNet etc.	3	4
10	Advances in the domain and Case Studies from various applications of NLP such as Question answering system, Information retrieval, Text categorization, text summarization, Named Entity Recognition, & Sentiment Analysis	4	4

TEXTBOOKS

1. Dan Jurafsky, James H. Martin, “Speech and Language Processing”, Stanford University, 2017
2. Christopher D.Manning and Hinrich Schuetze, “Foundations of Statistical Natural Language Processing”, MIT press, 1999.

RECOMMENDED READING

- 1 Joseph D. Booth, “Natural Language Processing”, Syncfusion, Inc., 2018
- 2 Shuly Wintner, “Formal Language Theory for Natural Language Processing”, ESSLLI, 2001
- 3 Steven Bird, Ewan Klein and Edward Loper, “Natural Language Processing with Python”, O'Reilly Media, 2009