

**III B.Tech I Semester**

23A31501	<b>NATURAL LANGUAGE PROCESSING (Professional Core)</b>	<b>L</b>	<b>T</b>	<b>P</b>	<b>C</b>
		<b>3</b>	<b>0</b>	<b>0</b>	<b>3</b>

**Course Objectives (COs)**

- Basics of NLP, Morphology, Tokenization, N-gram Models
- POS Tagging, Parsing, Treebanks, Ambiguity Handling
- Word Sense Disambiguation, Semantic Parsing, Sentiment Analysis
- Machine Translation, Transformers, BERT/GPT, Ethical NLP
- Speech Recognition, Feature Extraction, Discourse Analysis

**Course Outcomes**

- Understand morphological processing and the structure of words and documents.
- Analyze syntactic structures using various parsing algorithms.
- Apply semantic parsing techniques to interpret natural language text.
- Understand predicate-argument structures and meaning representation systems.
- Apply cross-lingual language models and speech recognition techniques in NLP applications

**UNIT I: Introduction to NLP**

Introduction to NLP: Origins and Challenges, Language and Grammar in NLP, Regular Expressions and Finite-State Automata, Tokenization: Text Segmentation and Sentence Splitting, Morphological Parsing: Stemming and Lemmatization, Spelling Error Detection and Correction, Minimum Edit Distance and Applications, Statistical Language Models: Unigram, Bigram, and Trigram Models, Processing Indian Languages in NLP.

**UNIT II: Word-Level and Syntactic Analysis**

Introduction, Part-of-Speech (POS) Tagging: Rule-Based, Stochastic and Transformation-Based Approaches, Hidden Markov Models (HMM) and Maximum Entropy Models for POS Tagging, Context-Free Grammar (CFG) and Constituency Parsing, Treebanks and Normal Forms for Grammar, Top-Down and Bottom-Up Parsing Strategies, CYK Parsing Algorithm, Probabilistic Context-Free Grammars (PCFGs), Feature Structures and Unification.

**UNIT III: Text Classification and Information Retrieval**

Naïve Bayes Classifier for Text Classification, Training and Optimization for Sentiment Analysis, Information Retrieval: Basic Concepts and Design Features, Information Retrieval Models: Classical, Non-Classical, and Alternative Models, Cluster Model, Fuzzy Model, and LSTM-Based Information Retrieval, Word Sense Disambiguation (WSD) Methods: Supervised and Dictionary-Based Approaches.

**UNIT IV: Machine Translation and Semantic Processing**

Introduction to Machine Translation (MT), Language Divergence and Typology in MT Encoder-Decoder Model for Machine Translation, Translating in Low-Resource Scenarios, MT Evaluation Metrics and Techniques, Bias and Ethical Issues in NLP and Machine Translation, Semantic Analysis and First-Order Logic in NLP, Thematic Roles and Selectional Restrictions in Semantics, Word Senses and Relations Between Senses

## **UNIT V: Speech Processing and Advanced NLP Models**

Speech Fundamentals: Phonetics and Acoustic Phonetics, Digital Signal Processing in Speech Analysis, Feature Extraction in Speech: Short-Time Fourier Transform (STFT), Mel-Frequency Cepstral Coefficients (MFCC) and Perceptual Linear Prediction (PLP), Hidden Markov Models (HMMs) in Speech Recognition.

### **Textbooks (Core Learning Materials)**

1. Daniel Jurafsky & James H. Martin – Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Pearson Education, 2023.
2. Tanveer Siddiqui & U.S. Tiwary – Natural Language Processing and Information Retrieval, Oxford University Press.

### **Reference Books (Supplementary Learning)**

1. T.V. Geetha – Understanding Natural Language Processing – Machine Learning and Deep Learning Perspectives, Pearson, 2024.
2. Akshay Kulkarni & Adarsha Shivananda – Natural Language Processing Recipes - Unlocking Text Data with Machine Learning and Deep Learning using Python, Apress, 2019.

### **Web links and Video Lectures (e-Resources):**

1. <https://www.youtube.com/watch?v=M7SWr5xObkA>
2. [https://onlinecourses.nptel.ac.in/noc23\\_cs45/preview](https://onlinecourses.nptel.ac.in/noc23_cs45/preview)
3. <https://archive.nptel.ac.in/courses/106/106/106106211/>