**NATURAL LANGUAGE PROCESSING LAB**

**19DS970**

**LAB MANUAL**

**2024-2025**

**ODD SEMESTER**

**List of Exercises**

**Week 1:**

* **Lexical/Morphological Analysis:**

Introduction to NLP tools and Libraries (NLTK, Spacy, Text blob)

Tokenization: Write a program to split text into words, punctuation marks, or even numbers.

* + Text Preprocessing: Develop functions for tasks like converting text to lowercase, removing special characters, or handling whitespace.

**Week 2:**

* **Syntactic Analysis:**

Language’s words and phrases arrange to form sentences. Syntactic analysis checks word arrangements for proper grammar

* + POS Tagging: Practice identifying parts-of-speech (POS) tags (e.g., noun, verb) for words in a sentence using Markov models and Viterbi algorithms
  + Syntax Parsing using Context Free Grammars

**Week 3:**

* **Semantic Analysis:**
  + Word Sense Disambiguation
  + Named Entity Recognition

**Week 4:**

* **Discourse Integration** (Hobbs’ Algorithm for Pronoun Resolution)
  + Text Summarization

**Week 5:**

* Text Representation
* BOW with TF-IDF
* N‐Gram Language Model
* Word2Vec
* Glove
* Fast Text
* Sentence embedding Technique: Doc2Vec
* Transformer – BERT

**Week 6:**

• Multi Page Document Classification

• Spam E-mail Detection

• Topic Modeling

• Content-based product recommendations

**Week 7:**

**• Pragmatic Analysis**

* + Speaker Intent Classification:
  + Context-aware Sentiment Analysis (Hate Speech Detection)

**Week 8 to 10:**

**Mini Project on NLP Applications (Not limited to this)**

* Question Answering (text Chat-bot)
* Large Language Modeling
* Machine Translation
* Spelling correction
* Auto text prediction (word level and sentence level)
* Speech Recognition