**CDS 3400 Natural Language Processing**

**Contact Hours 45**

Pre requisite Interactive programming in Python, Probability and statistics

**Purpose of the Course**

This course equips a learner with crucial knowledge, concepts and techniques of processing a natural language using computers.

**Expected Learning Outcomes**

By the end of the course learner should be able to:

1. Demonstrate knowledge of various techniques of natural language processing
2. Demonstrate knowledge of natural language modelling approaches
3. Apply Hidden Markov Models and POS Tagging methods to process a natural language
4. Extract morphemes from a natural language collection of text

**Course Outline**

| **Week** | **Topic** | **Subtopic** | **Assignment** | **Book Reference Chapter** |
| --- | --- | --- | --- | --- |
| 1 | Introduction to NLP | What is NLP? | Read Chapter 1 of "Speech and Language Processing" by Jurafsky and Martin | Chapter 1 |
| 2 | Text Preprocessing | Tokenization, Stop Words, Stemming, and Lemmatization | Implement a text preprocessing pipeline | Chapter 2 |
| 3 | Language Modeling | N-grams, Smoothing, and Perplexity | Build a language model for a given corpus | Chapter 3 |
| 4 | Part-of-Speech Tagging | Hidden Markov Models and Viterbi Algorithm | Implement a POS tagger for a given corpus | Chapter 8 |
| 5 | Named Entity Recognition | Rule-based and Machine Learning Approaches | Implement a named entity recognizer for a given corpus | Chapter 22 |
| 6 | Sentiment Analysis | Lexicon-based and Machine Learning Approaches | Implement a sentiment analyzer for a given corpus | Chapter 20 |
| 7 | Dependency Parsing | Dependency Trees and Transition-based Parsing | Implement a dependency parser for a given sentence | Chapter 14 |
| 8 | Semantic Role Labeling | FrameNet and PropBank | Implement a semantic role labeler for a given sentence | Chapter 15 |
| 9 | Discourse Analysis | Rhetorical Structure Theory and Coherence Relations | Analyze the discourse structure of a given text | Chapter 22 |
| 10 | Machine Translation | Rule-based and Statistical Approaches | Implement a machine translator for a given language pair | Chapter 25 |
| 11 | Question Answering | Information Retrieval and Knowledge-based Approaches | Implement a question answering system for a given corpus | Chapter 27 |
| 12 | Dialogue Systems | Task-oriented and Conversational Approaches | Implement a dialogue system for a given task | Chapter 24 |
| 13 | Text Generation | Rule-based and Neural Approaches | Generate text for a given prompt | Chapter 24 |
| 14 | Review and Future Directions | Recent Advances and Open Problems | Write a review paper on a recent NLP research paper | N/A |

**Mode of delivery**

Lectures

Practical and tutorial sessions in Computer Laboratory

Individual and group assignments

Class exercises and project work

**Instructional Materials**

Overhead projector and computer, handouts, white boards, appropriate software.

**Course Assessment**

Type Weighting (%)

Examination 70%

Continuous Assessment 30%

Total 100%

**Core Reading Materials**

The following are some reference books that can be used for this course:

1. "Speech and Language Processing" by Jurafsky and Martin
2. "Foundations of Statistical Natural Language Processing" by Manning and Schütze
3. "Natural Language Processing with Python" by Bird, Klein, and Loper
4. "Handbook of Natural Language Processing" edited by Indurkhya and Damerau