NATURAL LANGUAGE PROCESSING [3 0 0 3]

[Revised Credit System]

(Effective from the academic year 2022-2023)

SEMESTER – VI

Subject Code	**CSE XXXX**	**IA Marks**	**50**
Number of Lecture Hours/Week	**03**	**Exam Marks**	**50**
Total Number of Lecture Hours	**36**	**Exam Hours**	**03**

CREDITS – 03

Course objectives: This course will enable students to

- Get introduced to the concepts in Natural Language Processing
- Understand the variety of issues in processing various Natural Languages and mechanisms
- Familiarize with the applications of Natural Language Processing

Module -1	**Teaching Hours**
INTRODUCTION TO NATURAL LANGUAGE PROCESSING: Knowledge in Speech and Language Processing, Ambiguity, Models and Algorithm. **Text Book 1:** Chapter1: 1.1-1.3	**2 Hours**

Module -2

WORDS AND TRANSDUCERS: Survey of English Morphology, Finite-State Morphological Parsing, Building a Finite-State Lexicon, Finite-State Transducers, FSTs for Morphological Parsing, Lexicon-Free FSTs: Transducers and Orthographic rules, The combination of an FST and Lexicon rules, Words and sentence tokenization, Detecting and Correcting Spelling Errors, Minimum Edit Distance. Case study: Useful Applications of Regular Expressions, Normalizing Text, Regular Expression for tokenizing text, Segmentation **Text Book 1:** Chapter 3: 3.1-3.7, 3.9, 3.10 **Text Book 2:** Chapter 3: 3.6-3.8	**8 Hours**

Module – 3

N-GRAMS:	**6 Hours**

Counting words in Corpora, Simple(Unsmoothed) N-Grams, Training and Test Data, Smoothing, Interpolation, Backoff **Text Book 1:** Chapter 4: 4.1-4.3, 4.5-4.7	
Module-4	
WORD CLASSES AND PART OF SPEECH TAGGING: English Word Classes, Tag-sets for English, Part-of-Speech Tagging, Rule-based Part-of-Speech Tagging, HMM Part-of-Speech Tagging, Transformation-based Tagging, Evaluation and Error Analysis, Advance: The Noisy Channel Model for Spelling. Case study: Using a tagger, Tagged Corpora, Automatic Tagging, N-Gram Tagging. **Text Book 1:** Chapter 5: 5.1-5.7, 5.9 **Text Book 2:** Chapter 5: 5.1, 5.2, 5.4-5.7	**8 Hours**
Module-5	
FORMAL GRAMMARS OF ENGLISH: Constituency, Some Grammar Rules for English, Treebanks: Example of The Penn Treebank project, Dependency Grammar. **Text Book 1:** Chapter 12: 12.1-12.3, 12.4.1, 12.7	**5 Hours**
Module-6	
PARSING WITH CONTEXT FREE GRAMMARS: Parsing as Search, Ambiguity, and Dynamic Programming Parsing Method: CKY algorithm. **Text Book 1 :**Chapter 13: 13.1, 13.2, 13.4.1	**5 Hours**
Module-7	
STATISTICAL PARSING: Probabilistic Context-Free Grammars, Evaluating Parsers. **Text Book 1:** Chapter 14: 14.1, 14.7	**2 Hours**
Course outcomes:	
After studying this course, students will be able to: 1. Describe the basic concepts and techniques of Natural Language Processing 2. Describe POS tagging and Context-Free Grammar of English language for Processing	

3. Explain the elements and applications of parsing and semantic analysis.
4. Describe the basics of knowledge representation for Natural Language Processing applications
5. Assess the relationship between, Natural Language Processing, Machine Learning and Statistics.

Text Books:

1. Daniel Jurafsky & James H. Martin, *Speech and Language Processing*, (2e), Pearson, 2009.
2. Steven Bird, Ewan Klein and Edward Loper, *Natural Language Processing with Python*, (1e), O'Reilly Media, 2009

Reference Books:

1. Akshar Bharati, Rajeev Sangal and Vineet Chaitanya, *Natural Language Processing: A Paninian Perspective* , Prentice-Hall of India, New Delhi, 1995.
2. Steven Bird, Ewan Klein, Edward Loper, *Natural Language Processing with Python – Analysing Text with natural language toolkit* , O'Reilly Media, 2009.
3. Chris Manning, Hinrich Schutze , *Foundations of Statistical Natural Language Processing*, MIT Press, Cambridge, 1999.