No:-   Date:

CSX4226:   Natural Language Processing Techniques

L–T–P-Cr: 2-0-2-3

Overview:
This course introduces the theory and methods of natural language processing (NLP). NLP systems understand and produce human language for applications such as information extraction, machine translation, automatic summarization, question-answering, and interactive dialog systems. The course covers knowledge-based and statistical approaches to language processing for syntax (language structures), semantics (language meaning), and pragmatics/discourse (the interpretation of language in context) with various applications.

Pre-requisite: Basic knowledge of compiler and automata theory.

Objectives:

 This subject aims to achieve the following goals:

- To introduce students, the challenges of empirical methods for natural language processing (NLP) applications.
- To introduce basic mathematical models and methods used in NLP applications to formulate computational solutions.
- To provide students with the knowledge on designing procedures for natural language resource annotation and the use of related tools for text analysis and hands-on experience of using such tools.
- To introduce students research and development work in information retrieval, information extraction, and knowledge discovery using different natural language resources.
- To give an overview of the major technologies in speech recognition and synthesis including tools for acoustic analysis and hands-on experience of using such tools.
- To give students opportunities to sharpen their programming skills for computational linguistics applications

Course outcomes:

Upon completion of the subject, students will be able to:

No.	Course outcomes	Mapping to POs
1	Understanding of the fundamental mathematical models and algorithms in the field of NLP.	PO1, PO3
2	Apply these mathematical models and algorithms in applications in software design and implementation for NLP.	PO1, PO3
3	Understand the principles of language resource annotation and its use in machine learning applications and apply the above principles in analysis of data and acquire intended information through the use of available tools.	PO1, PO4-PO7, PO12
4	Understand the design and implementation issues in various NLP applications such as information retrieval and information extraction.	PO4, PO5
5	Understand the complexity of speech and the challenges facing speech engineers.	PO3, PO4-PO7
6	Understand the principles of automatic speech recognition and synthesis.	PO1, PO4-PO7, PO12

UNIT I   **Lectures: 2**

Introduction: Need for Processing Natural languages, Issues in NLP and Complexity of Processing NLP, Brief history of NLP application development

UNIT II   **Lectures: 14**

Regular Expressions, Text Normalization, and Edit Distance, Language Modeling with N-Grams, Naive Bayes Classification and Sentiment, Logistic Regression, Vector Semantics, Neural Nets and Neural Language Models, Part-of-Speech Tagging, Sequence Processing with Recurrent Networks

UNIT III   **Lectures: 14**

Formal Grammars of English, Syntactic Parsing, Statistical Parsing, Dependency Parsing, The Representation of Sentence Meaning, Computational Semantics, Semantic Parsing, Information Extraction, Semantic Role Labeling and Argument Structure, Lexicons for Sentiment, Affect, and Connotation, Co-reference Resolution and Entity Linking, Discourse Coherence

UNIT IV   **Lectures: 12**

Machine Translation, Question Answering, Dialog Systems and Chatbots, Advanced Dialog Systems, Speech Recognition and Synthesis. Applications: Question-Answering System, Text Summarizers, Applications of NLP.

Reference Books:

1) Dan Jurafsky and James H. Martin, *Speech and Language processing*, 2nd eddition, Prentice Hall, 2008
2) Christopher Manning and Hinrich Schuetze, *Foundations of Statistical Natural Language Processing*, the MIT Press, 1999