III B.TECH II SEM	L	T	P	C
	3	0	0	3
PE1: Natural LanguageProcessing				

Course Objectives:

- This course introduces the fundamental concepts and techniques of natural language processing (NLP).
- Students will gain an in-depth understanding of the computational properties of natural languages and the commonly used algorithms for processing linguistic information.
- The course examines NLP models and algorithms using both the traditionalsymbolic and the more recent statistical approaches.
- Enable students to be capable to describe the application based on natural language processing and to show the points of syntactic, semantic and
- pragmatic processing.

UNIT - I:

Introduction: What is Natural Language Processing (NLP), Origins of NLP, Language and Knowledge, The challenges of NLP, Language and Grammar, Processing Indian Languages, NLP Applications, Some successful Early NLP Systems, Information Retrieval, **Language Modelling:** Introduction, Various Grammar-based Language Models, Statistical Language Model.

UNIT - II:

Word Level Analysis: Introduction, Regular Expressions, Finite State Automata, Morphological Parsing, Spelling Error Detection and Correction, Words and Word Classes, Part-of-Speech Tagging, **Syntactic Analysis:** Introduction, Context- Free Grammar, Constituency, Parsing, Probabilistic Parsing, Indian Languages.

UNIT - III:

Semantic Analysis and Pragmatics: Introduction, Meaning Representation, Lexical Semantics, Ambiguity, Word Sense Disambiguation, **Discourse Processing:** Introduction, Cohesion, Reference Resolution, Discourse Coherence and Structure.

UNIT - IV:

Natural Language Generation: Introduction, Architectures of NLG Systems, Generation task and Representations, Applications of NLG,

Machine Translation: Introduction, Problems in Machine Translation, Characteristics of Indian Languages, Machine Translation Approaches, Direct Machine Translation, Rule-based Machine Translation, Corpus-based Machine Translation, Semantic or Knowledge-based MT Systems, Translation involving Indian Languages.

UNIT - V:

NLP Applications: Introduction, Information Extraction, Automatic Text Summarization, Question-Answering System, **Lexical Resources:** Introduction, Word Net, Frame Net, Stemmers, Part-of-Speech Tagger, Research Corpora, Journals and Conferences in the Area.

TEXT BOOKS:

1. Tanveer Siddiqui, U.S. Tiwary, ―Natural Language Processing and Information Retrieval, Oxford University Press, 2008.
2. Daniel Jurafsky, James H. Martin―Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech, Pearson Publication, 2014.

REFERENCE BOOKS:

1. Steven Bird, Ewan Klein and Edward Loper, ―Natural Language Processingwith Python, First Edition, OReilly Media, 2009.
2. Breck Baldwin, ―Language processing with Java and Ling Pipe Cookbook, Atlantic Publisher, 2015.
3. Richard M Reese, ―Natural Language Processing with Java, OReilly Media,2015.

Course Outcomes:

- Demonstrate a given text with basic Language features.
- Explain a rule based system to tackle morphology/syntax of a language.
- To design an innovative application using NLP components. K6 To design a tagset to be used for statistical processing for real-time applications.
- To compare and contrast the use of different statistical approaches for differenttypes of NLP applications.