

PCS6J003|NATURAL LANGUAGE PROCESSING (4-0-0)

MODULE - I : (8 HOURS)

Introduction to Natural Language Processing: Need for processing natural languages, issues and processing complexities in NLP, Brief history of NLP application development, Overview of phases of natural language processing; Language Modeling: Various types of languages and its modelling, Grammar based language models, Government and binding, Lexical functional grammar for handling natural languages, Statistical modelling, n gram model.

MODULE -II : (12 HOURS)

Word Level Analysis: Regular expressions, Finite state automata Morphological parsing, Human morphological parsing, Spelling error detection and correction, Words and word classes, Part of speech tagging, Rule-based tagger, Stochastic tagger, Unknown words. Syntactic Analysis: Context free grammar, Phrase and sentence level constructions, Parsing: Top-down parsing, Bottom-up parsing, A basic top-down parser, The Earley parser, The CYK Parser, Probabilistic parsing: Estimating rule probabilities, Parsing PCFGs, Problems with PCFG.

MODULE - III : (12 HOURS)

Semantic Analysis: Meaning representation, Characteristics of meaning representation languages, Meaning structure of languages, Syntax-driven semantic analysis, Semantic grammars, Lexical Semantics, Relations among lexemes and their senses, WordNet, Internal structure of words, Ambiguity, Word sense disambiguation, Selectional restriction-based disambiguation, Robust word sense disambiguation, Information retrieval, Other Information retrieval tasks

MODULE - IV : (12 HOURS)

Discourse Analysis: Context-based word sense disambiguation approaches, Knowledge sources in WSD, WSD evaluation discourse context and world knowledge: Local discourse context, Anaphora resolution, World Knowledge, Discourse structure, Psycholinguistic studies of reference and coherence, Natural Language Generation: Architecture of language generators, Surface realization, Discourse planning, Template-based, phrase-based and feature-based natural language generation, Knowledge-based approaches.

MODULE - V : (12 HOURS)

Machine Translation: Machine translation system, Issues, Language similarities and differences, The transfer metaphor, The interlingua idea: using meaning, Machine translation approaches: Direct machine translation, Rule-based machine translation: Transfer-based and interlingua based, Statistical and example-based machine translation, Semantic or knowledge based machine translation systems. Advanced Applications of NLP: Text to Speech system, Speech to Text system, Question Answering System, Text Summarization.

Text Books:

1. Speech and Language Processing – An introduction to language processing, computational linguistics, and speech recognition, D. Jurafsky and J. H. Martin, 2003, Pearson Education.
2. Natural language Processing and Information Retrieval, T. Siddiqui and U. S. Tiwary, 2008, Oxford University Press.

Reference Books:

1. Natural Language Understanding, J. Allen, 2nd edition, 2008, Pearson Education.
2. Natural Language Processing- A Pananian Perspective, A. Bharathi, V. Chaitanya and R. Sangal, 1995, PHI Learning.
3. Foundations of Statistical Natural Language Processing, Christopher D. Manning and Hinrich Schütze, 1999, MIT Press.