

Course Code	Course Name	Credits
DLO8012	Natural Language Processing	4

Course objectives:

1. To understand natural language processing and to learn how to apply basic algorithms in this field.
2. To get acquainted with the basic concepts and algorithmic description of the main language levels: morphology, syntax, semantics, and pragmatics.
3. To design and implement applications based on natural language processing
4. To implement various language Models.
5. To design systems that uses NLP techniques

Course outcomes: On successful completion of course learner should:

1. Have a broad understanding of the field of natural language processing.
2. Have a sense of the capabilities and limitations of current natural language technologies,
3. Be able to model linguistic phenomena with formal grammars.
4. Be able to Design, implement and test algorithms for NLP problems
5. Understand the mathematical and linguistic foundations underlying approaches to the various areas in NLP
6. Be able to apply NLP techniques to design real world NLP applications such as machine translation, text categorization, text summarization, information extraction...etc.

Prerequisite: Data structure & Algorithms, Theory of computer science, Probability Theory.

Module No.	Unit No.	Topics	Hrs.
1	Introduction	History of NLP, Generic NLP system, levels of NLP , Knowledge in language processing , Ambiguity in Natural language , stages in NLP, challenges of NLP ,Applications of NLP	4
2	Word Level Analysis	Morphology analysis –survey of English Morphology, Inflectional morphology & Derivational morphology, Lemmatization, Regular expression, finite automata, finite state transducers (FST) ,Morphological parsing with FST , Lexicon free FST Porter stemmer. N –Grams- N-gram language model, N-gram for spelling correction.	10
3	Syntax analysis	Part-Of-Speech tagging(POS)- Tag set for English (Penn Treebank) , Rule based POS tagging, Stochastic POS tagging, Issues –Multiple tags & words, Unknown words. Introduction to CFG, Sequence labeling: Hidden Markov Model (HMM), Maximum Entropy, and Conditional Random Field (CRF).	10
4	Semantic Analysis	Lexical Semantics, Attachment for fragment of English- sentences, noun phrases, Verb phrases, prepositional phrases, Relations among lexemes & their senses –Homonymy, Polysemy, Synonymy, Hyponymy, WordNet, Robust Word Sense Disambiguation (WSD) ,Dictionary based approach	10

5	Pragmatics	Discourse –reference resolution, reference phenomenon , syntactic & semantic constraints on co reference	8
6	Applications (preferably for Indian regional languages)	Machine translation, Information retrieval, Question answers system, categorization, summarization, sentiment analysis, Named Entity Recognition.	10