Course Code	Course Name	Credit
CSDC7013	**Natural Language Processing**	**03**

Pre-requisite: Theory of Computer Science, System Programming &Compiler Construction	

	Course Objectives: The course aims
1	To define natural language processing and to learn various stages of natural language processing.
2	To describe basic concepts and algorithmic description of the main language levels: Morphology, Syntax, Semantics, and Pragmatics &Discourse analysis.
3	To design and implement various language models and POS tagging techniques.
4	To design and learn NLP applications such as Information Extraction, Question answering.
5	To design and implement applications based on natural language processing.

	Course Outcomes: Students will be able
1	To describe the field of natural language processing.
2	To design language model for word level analysis for text processing.
3	To design various POS tagging techniques and parsers.
4	To design, implement and test algorithms for semantic and pragmatic analysis.
5	To formulate the discourse segmentation and anaphora resolution.
6	To apply NLP techniques to design real world NLP applications.

Module		Detailed Content	Hours
1	1.1	**Introduction to NLP**	3
		Origin & History of NLP; Language, Knowledge and Grammar in language processing; Stages in NLP;Ambiguities and its types in English and Indian Regional Llanguages; Challenges of NLP;Applications of NLP	
	1.2	**Self-Learning topics:** Variety types of tools for regional languages pre-processing and other functionalities	
2	2.1	**Word Level Analysis**	9
		Basic Terms: Tokenization, Stemming, Lemmatization; Survey of English Morphology, Inflectional Morphology, Derivational Morphology; Regular expression with types; Morphological Models: Dictionary lookup, finite state morphology; Morphological parsing with FST (Finite State Transducer);Lexicon free FST Porter Stemmer algorithm; Grams and its variation: Bigram, Trigram; Simple (Unsmoothed) N-grams; N-gram Sensitivity to the Training Corpus; Unknown Words: Open versus closed vocabulary tasks; Evaluating N-grams: Perplexity;	

		Smoothing: Laplace Smoothing, Good-Turing Discounting;	
	2.2	**Self-Learning topics:** Noisy channel models, various edit distance, Advance Issues in Language Modelling	
3	3.1	**Syntax analysis**	10
		Part-Of-Speech tagging(POS); Tag set for English (Upenn Treebank); Difficulties /Challenges in POS tagging; Rule-based, Stochastic and Transformation-based tagging; Generative Model: Hidden Markov Model (HMM Viterbi) for POS tagging; Issues in HMM POS tagging; Discriminative Model: Maximum Entropy model, Conditional random Field (CRF);Parsers: Top down and Bottom up; Modelling constituency; Bottom Up Parser: CYK, PCFG (Probabilistic Context Free Grammar), Shift Reduce Parser; Top Down Parser: Early Parser, Predictive Parser	
	3.2	**Self-Learning topics:** Evaluating parsers, Parsers based language modelling, Regional languages POS tree banks	
4	4.1	**Semantic Analysis**	7
		Introduction, meaning representation; Lexical Semantics; Corpus study; Study of Various language dictionaries like WorldNet, Babelnet; Relations among lexemes & their senses –Homonymy, Polysemy, Synonymy, Hyponymy; Semantic Ambiguity; Word Sense Disambiguation (WSD); Knowledge based approach(Lesk's Algorithm), Supervised (Naïve Bayes, Decision List),Introduction to Semi-supervised method (Yarowsky) Unsupervised (Hyperlex)	
	4.2	**Self-Learning topics:** Dictionaries for regional languages, Distributional Semantics, Topic Models	
5	5.1	**Pragmatic & Discourse Processing**	5
		Discourse: Reference Resolution, Reference Phenomena, Syntactic & Semantic constraint on coherence; Anaphora Resolution using Hobbs and Cantering Algorithm	
	5.2	**Self-Learning topics:** Discourse segmentation, Conference resolution	
6	6.1	**Applications of NLP**	5
		Case studies on (preferable in regional language):Machine translation; Text Summarization; Sentiment analysis; Information retrieval; Question Answering system	
	6.2	**Self-Learning topics:** Applications based on Deep Neural Network with NLP such as LSTM network, Recurrent Neural network etc.	

Textbooks:	
1	Daniel Jurafsky, James H. and Martin, Speech and Language Processing, Second Edition, Prentice Hall, 2008.
2	Christopher D.Manning and HinrichSchutze, Foundations of Statistical Natural Language Processing, MIT Press, 1999.
References:	
1	Siddiqui and Tiwary U.S., Natural Language Processing and Information Retrieval, Oxford University Press, 2008.
2	Daniel M Bikel and ImedZitouni ― Multilingual natural language processing applications: