Sardar Patel Institute of Technology

Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India
(Autonomous Institute Affiliated to University of Mumbai)

Course Code	Course Name	Teaching Scheme (Hrs/week)			Credits Assigned			
		L	T	P	L	T	P	Total
CEE91C	Natural Language Processing(NLP)	3	--	--	3	--	--	3
		Examination Scheme						
		ISE	MSE	ESE				
		10	30	100 (60% Weightage)				

Pre-requisite Course Codes	Programming Methodology & Data Structure Probability & Statistics Theory of Computer Science
At the end of successful completion of the course, students will be able to	
	CO1	To model linguistic phenomena with formal grammars.
Course Outcomes	CO2	To design, implement, and analyze NLP algorithms.
	CO3	Apply NLP techniques to design real world NLP applications.
	CO4	Implement proper experimental methodology for training and evaluating empirical NLP systems.

Module No.	Unit No.	Topics	Ref.	Hrs.
1		Introduction	1,4	3
	1.1	History of NLP, Generic NLP system, levels of NLP , Knowledge in language processing ,		
	1.2	Ambiguity in Natural language , stages in NLP, challenges of NLP ,Applications of NLP- Machine translation,		
	1.3	question answering system, Information retrieval, Text categorization , text summarization & Sentiment Analysis		
2		Word Level Analysis	1,3,4	7
	2.1	Morphology analysis –survey of English Morphology, Inflectional morphology & Derivational morphology;		
	2.2	Regular expression, finite automata, finite state transducers (FST) ,Morphological parsing with FST ,		
	2.3	Lexicon free FST - Porter stemmer. N –Grams- N-gram language model, Ngram for spelling correction.		
3		Syntax analysis	1,3	8
	3.1	Part-Of-Speech tagging(POS)- Tag set for English (Penn Treebank) , Rule based POS tagging,		
	3.2	Stochastic POS tagging, Issues –Multiple tags & words, Unknown words, class based n –grams.		
	3.3	Context Free Grammar – Constituency , Context free rules & trees, Sentence level construction , Noun Phrase, coordination,		

Sardar Patel Institute of Technology

Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India
(Autonomous Institute Affiliated to University of Mumbai)

		agreement, the verb phrase & sub categorization		
4		Semantic Analysis		8
	4.1	Attachment for fragment of English- sentences, noun phrases, Verb phrases, prepositional phrases,	1,2,3	
	4.2	Relations among lexemes & their senses –Homonymy, Polysemy, Synonymy, Hyponymy, Wordnet,		
	4.3	Selectional restriction based disambiguation & limitations , Robust WSD – machine learning approach and dictionary based approach		
5		Pragmatics	1,2,3	8
	5.1	Discourse –reference resolution, reference phenomenon , syntactic & semantic constraints on co reference,		
	5.2	preferences in pronoun interpretation , algorithm for pronoun resolution .Text coherence, discourse structure		
6		Applications (preferably for Indian regional languages)	1,2,3,4,5	8
	6.1	Machine translation, Information retrieval.		
	6.2	Question answers system, categorization, summarization, sentiment analysis		
			Total	**42**

References:

[1] Daniel Jurafsky, James H. Martin "Speech and Language Processing" Second Edition, Prentice Hall, 2008.

[2] Christopher D.Manning and Hinrich Schutze, "Foundations of Statistical Natural Language Processing ", MIT Press, 1999.

[3] Siddiqui and Tiwary U.S., Natural Language Processing and Information Retrieval, Oxford University Press (2008).

[4] Daniel M Bikel and Imed Zitouni " Multilingual natural language processing applications" Pearson, 2013

[5] Alexander Clark (Editor), Chris Fox (Editor), Shalom Lappin (Editor) " The Handbook of Computational Linguistics and Natural Language Processing " ISBN: 978-1-118-