

Branch Name:	Artificial Intelligence and Data Science
Program Code:	ET108
Semester/Year:	Semester VII/ Fourth Year
Course Name:	Professional Elective –IV: Natural Language Processing
Course Code:	2ET1080703T
Pre-requisite Course:	Probability and statistics, Programming and data structures.

Course Objectives:

1. The Natural Language Processing course gives a detailed look at the science of applying machine learning algorithms to process large amounts of natural language data.
2. NLP is driving the growth of the AI market, and this course helps to develop the skills required to become an NLP Engineer.

Teaching and Examination Scheme:

Credit	Teaching Scheme Hours per week				Evaluation Scheme (Marks)								
					Theory (Marks)				Practical (Marks)				Total (Marks)
	L	T	P	Total	SEE		CCA		SEE		CCA		
					Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	
3	3	-	-	3	24	60	16	40	-	-	-	-	100

Subject Contents:

Sr. No	Topic	Total Hours	Weightage (%)
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 Self learning topics: Empirical laws	6	10
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	9	20
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 Self learning topics: Conditional Random Field (CRF)	9	20
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, Robust Word Sense Disambiguation (WSD),Dictionary based approach	10	25

5	Text Summarization , Text Classification Text summarization- LEXRANK, Optimization based approaches for summarization , Summarization evaluation, Text classification	6	15
6	Sentiment Analysis and Opinion Mining Sentiment Analysis introduction, Sentiment Analysis - Affective lexicons, Learning affective lexicons, Computing with affective lexicons, Aspect based sentiment analysis	5	10

List of Reference Books:

1. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition Jurafsky, David, and James H. Martin, PEARSON 2000
2. Foundations of Statistical Natural Language Processing, Manning, Christopher D., and Hinrich Schütze, Cambridge, MA: MIT Press 1999.
3. Natural Language Understanding, James Allen. The Benjamin/Cummings Publishing Company Inc.1988.
4. Natural Language Processing with Python by Steven Bird, Ewan Klein, Edward Loper
Released June 2009 Publisher(s): O'Reilly Media, Inc.

Course Learning Outcomes (CLO): On completion of this course, the students will be able to:

CLO	Description	Bloom's Taxonomy Level
CLO1	Understand comprehend the key concepts of NLP and Identify the NLP challenges and issues.	1 Remembering 2 Understanding
CLO2	Develop Language Modeling for various text corpora across the different languages	4 Analyze 5 Evaluate
CLO3	Illustrate computational methods to understand language phenomena of word sense disambiguation	4 Analyze
CLO4	Design and develop applications for text or information extraction/summarization/classification	2 Understanding 4 Analyze
CLO5	Apply different Machine translation techniques for translating a source to target language(s)	3 Applying 6 Creating

Mapping of CLOs with POs & PSOs

Course Learning Outcomes	Program Outcomes (POs)												Program Specific Outcomes(PSOs)			
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3	PSO4
CLO1			M		M						M		H	M	L	
CLO2	M	H	M		H						M	M	M	M	L	
CLO3		M	H	M	M							M	H	H	M	L
CLO4	L	L	M	M	M							M	M		M	H
CLO5	H	H	H	H	H						H					

H: High, M: Medium, L: Low