

Branch Name:	Computer Engineering (Software Engineering)
Program Code:	ET203
Course Title:	Natural Language Processing
Course Code:	2ET2030307T
Pre-requisite Course:	Limit of a function, Continuity and Differentiability of a function, Differentiation and Integration.

Course Objectives: After successful completion of the course, student will be able to

1. Understand NLP basic preliminaries.
2. Express in mathematical formulation.
3. Apply Differential and Integral Calculus in formal representation of various computing constructs.
4. Recognize the importance of NLP for analysis in engineering problems.

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

Course Contents:

Unit No	Topics	Total Hours	Weightage (%)
1	Introduction: What is NLP, Fundamental and Scientific goals, Engineering goals, stages of NLP, problems in NLP, Applications of NLP, Empirical Laws of language, zipf's law, Heap's law.	07	15
2	Text Processing: Tokenization, word token, word type, sentence segmentation, feature extraction, issues in tokenization for different languages, word segmentation, text segmentation, normalization, case folding, Spelling Correction, Morphology, Stemming, Porters Algorithm, , lemmatization, spelling correction - dynamic programming approach for finding edit distance, N-gram Language Modeling- context sensitive spelling correction, probabilistic language model, auto completion prediction, Evaluation and perplexity, Smoothing techniques.	08	20
3	Tagging of POS: Sequence labeling tasks of NLP, POS tagging, POS tag sets, Hidden Markov ModelIntroduction, Markov Processes, HMM characterization -Likelihood of a sequence (Forward Procedure, Backward Procedure), Best state sequence (Viterbi Algorithm), Re-estimation(BaumWelch - Forward-Backward Algorithm) , Models for Sequential tagging – Maximum Entropy, Conditional Random Field.	08	20
4	Importance of Syntax: Constituency and dependency parsing, Constituency parser -Syntactic structure, Parsing methodology, Different parsing algorithms, Parsing in case of ambiguity, Probabilistic parsing, CKY algorithm, Issues in parsing, Dependency parsing- Syntactic structure, Parsing methodology, Transition-Based Dependency Parsing, Graph-Based dependency parsing, Evaluation, Co-reference resolution, Named-entity recognition.	10	25
5	Semantic and knowledge base: WordNet: Word Senses, Word relations, Word similarity and thesaurus methods, Word sense disambiguation, WordNet. Lexical and Distributional Semantics - Introduction, models of semantics, applications	6	10

6	Word Embeddings: Introduction, one-hot vectors, methods of generating word embeddings, Skip-gram, CBOW, Glove model, Fast Text model, evaluation measures-rough scores.	6	10
---	--	---	----

Reference Books:

1. Daniel Jurafsky and James H. Martin, "Speech and Language Processing", Second Edition, Prentice Hall, 2008, ISBN: 978-0131873216.
2. Allen James, "Natural Language Understanding", Second Edition, Benjamin/Cumming, 1994, ISBN: 978-0805303346.
3. Chris Manning and Hinrich Schuetze, "Foundations of Statistical Natural Language Processing", MIT Press, ISBN: 978-0262133609.
4. Journals: Computational Linguistics, Natural Language Engineering, Machine Learning, Machine Translation, Artificial Intelligence.
5. Conferences: Annual Meeting of the Association of Computational Linguistics (ACL), Computational Linguistics (COLING), European ACL (EACL), Empirical Methods in NLP (EMNLP), Annual Meeting of the Special Interest Group in Information Retrieval (SIGIR), Human Language Technology (HLT).

Course Learning Outcomes (CLO):

Course Outcomes	Description	Bloom's Taxonomy Level
COL1	Demonstrate the understanding of basic text processing techniques in NLP	Understanding (2) Applying (3)
COL2	Design, implement and evaluate part-of-speech taggers and parsers for a language.	Applying (3)
COL3	To understand the Importance of Syntax in NLP	Understanding (2) Applying (3)
COL4	To Study and demonstrate the semantic and knowledge base	Applying (3)
COL5	Analyze and build word embeddings for different languages.	Applying (3)

Mapping of CLO with PO and PSO

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	H	M	M		M											
CLO2	H	M			M							L				
CLO3	H	M	M		M											
CLO4	M	M	M		M							L				
CLO5	M	M										L				

H: High, M: Medium, L: Low