

## NATURAL LANGUAGE PROCESSING

### (Professional Elective –IV)

<b>Course Code</b>	20IT4702D	<b>Year</b>	IV	<b>Semester</b>	I
<b>Course Category</b>	PE - IV	<b>Branch</b>	IT	<b>Course Type</b>	Theory
<b>Credits</b>	3	<b>L-T-P</b>	3-0-0	<b>Prerequisites</b>	Data Mining
<b>Continuous Internal Evaluation :</b>	30	<b>Semester End Evaluation:</b>	70	<b>Total Marks:</b>	100

Course Outcomes		Blooms Taxonomy Level
<b>Upon Successful completion of course, the student will be able to</b>		
<b>CO1</b>	Understand the theoretical foundations of natural language processing in linguistics and formal language theory.	L2
<b>CO2</b>	Apply algorithms to solve text categorization tasks.	L3
<b>CO3</b>	Use concepts of semantic and syntactic analysis in real world NLP applications.	L3
<b>CO4</b>	Analyze NLP tasks using existing algorithms and frameworks for various applications.	L4

<b>Contribution of Course Outcomes towards achievement of Program Outcomes &amp; Strength of Correlations (H:High, M:Medium, L:Low)</b>														
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO1	3												3	
CO2		3			3								3	
CO3			3											3
CO4	3													3

Syllabus		
Unit No	Contents	Mapped CO
<b>I</b>	<b>Regular Expressions, Text Normalization, Edit Distance-</b> Regular Expression, Words, Corpora, Text Normalization, Minimum Edit Distance. N-Gram Language Models-NGrams, Evaluating Language Models, Generalization and Zeros, Smoothing, Kneser-Ney Smoothing, The web and stupid Backoff, Advanced Perplexity's Relation to Entropy.	<b>CO1</b> <b>CO4</b>
<b>II</b>	<b>Naive Bayes and Sentiment Classification:</b> Naive Bayes Classifiers, Training the Naive Bayes Classifier, Worked example, Optimizing for Sentiment Analysis, Naive Bayes for other text classification tasks, Naive Bayes as a Language Model, Evaluation: Precision, Recall, F-measure, Test sets and Cross-validation, Statistical Significance Testing, Avoiding Harms in Classification	<b>CO1</b> <b>CO2</b> <b>CO4</b>
<b>III</b>	<b>Vector Semantics and Embeddings-</b> Lexical Semantics, Vector Semantics, Words and Vectors, Cosine for measuring similarity, TF-IDF: Weighing terms in the vector, Applications of the TF-IDF vector model, Word2vec, Visualizing Embeddings, Semantic properties of embeddings, Bias and Embeddings, Evaluating Vector Models.	<b>CO1</b> <b>CO3</b> <b>CO4</b>

IV	<b>Sequence Labeling for Parts of Speech and Named Entities-</b> English Word Classes, Part-of-Speech Tagging, Named Entities and Named Entity Tagging, HMM Part-of-Speech Tagging, Conditional Random Fields (CRFs), Evaluation of Named Entity Recognition	CO1 CO3 CO4
V	<b>Applications of NLP-</b> Question Answering Information Retrieval IR-based Factoid Question Answering, Entity Linking, Knowledge-based Question Answering, Using Language Models to do QA, Classic QA Models, Evaluation of Factoid Answers, Chatbots & Dialogue Systems, Properties of Human Conversation, Chatbots , GUS: Simple Frame-based Dialogue Systems, The Dialogue-State Architecture, Evaluating Dialogue Systems, Dialogue System Design	CO1 CO2 CO3 CO4

<b>Learning Resources</b>	
<b>Text Books</b>	
<ol style="list-style-type: none"> <li>1. Speech and Language Processing: An introduction to Natural Language Processing, Computational Linguistics and Speech Recognition by Daniel Jurafsky and James H Martin, 3<sup>rd</sup> Edition, Prentice Hall, 2020.</li> <li>2. Natural Language Processing: An information Access Perspective by Kavi Narayana Murthy, Ess Publications, 2006.</li> </ol>	
<b>References</b>	
<ol style="list-style-type: none"> <li>1. Applied Text Analysis with Python by Benjamin Bengfort, Tony Ojeda, Rebecca Bilbro, O'Reilly Media, June 2018.</li> <li>2. Natural Language Processing Recipes by Akshay Kulkarni, Adarsha Shivananda, Apress, 2019</li> </ol>	
<b>E-Resources and other Digital Material</b>	
<ol style="list-style-type: none"> <li>1. Natural Language Processing by Pawan Goyal, IIT Kharagpur, <a href="https://swayam.gov.in/nd1_noc19_cs56/preview">https://swayam.gov.in/nd1_noc19_cs56/preview</a></li> <li>2. Natural Language Processing offered by deeplearning.ai on Coursera <a href="https://www.coursera.org/specializations/natural-language-processing">https://www.coursera.org/specializations/natural-language-processing</a></li> </ol>	