NATURAL LANGUAGE PROCESSING

Course Code	19IT4701E	Year	IV	Semester	I
Course Category	PE	Branch	IT	CourseType	Theory
Credits	3	L-T-P	3-0-0	Prerequisites	Data Mining
Continuous Internal	30	Semester End Evaluation:	70	Total Marks:	100
Evaluation:					

(Program Elective-IV)

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

Contribution of Course Outcomes towards achievement of Program Outcomes & Strength of correlations (H:High, M:Medium, L:Low)														
	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					
Ι	Regular Expressions, Text Normalization, Edit Distance- 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.	CO1 CO4					
п	Naive Bayes and Sentiment Classification: 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	CO1 CO2 CO4					
ш	Vector Semantics and Embeddings- 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.	CO1 CO3 CO4					
IV	Sequence Labeling for Parts of Speech and Named Entities- 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	Applications of NLP- 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					

Learning Resources

TextBooks

- 1. Speech and Language Processing: An introduction to Natural Language Processing, Computational Linguistics and Speech Recognition by Daniel Jurafsky and James HMartin,3rd Edition,PrenticeHall,2019.
- 2. Natural Language Processing: An information Access Perspective by Kavi Narayana Murthy, Ess Publications, 2006.

References

- 1. Applied Text Analysis with Python by Benjamin Bengfort, Tony Ojeda, Rebecca Bilbro, O'Reilly Media, June 2018.
- 2. Natural Language Processing Recipes by Akshay Kulkarni, Adarsha Shivananda, Apress,2019
- E-Resources and other Digital Material
- 1. Natural Language Processing by Pawan Goyal, IIT Kharagpur, https://swayam.gov.in/nd1_noc19_cs56/preview
- 2. Natural Language Processing offered by deeplearning.ai on Coursera https://www.coursera.org/specializations/natural-language-processing