Course Code	Course Title	L	T	P	C
BCSE409L	Natural Language Processing	3	0	0	3
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. To introduce the fundamental concepts and techniques of Natural language Processing for analyzing words based on Morphology and CORPUS.
2. To examine the NLP models and interpret algorithms for classification of NLP sentences by using both the traditional, symbolic and the more recent statistical approach.
3. To get acquainted with the algorithmic description of the main language levels that includes morphology, syntax, semantics, and pragmatics for information retrieval and machine translation applications.

Course Outcomes

1. Understand the fundamental concepts of Natural Language Processing.
2. Develop useful systems for language processing and related tasks involving text processing and demonstrate text-based processing of natural language with respect to morphology.
3. Check the syntactic and semantic correctness of natural language.
4. Select a suitable language modelling & Feature Representation to develop real-world applications.
5. Develop computational methods for real-world applications using deep learning.

Module:1	Introduction to NLP	4 hours

Introduction to NLP – Ambiguous nature of NLP - Morphological Analysis - Syntax Analysis– Semantic Analysis –Pragmatic Analysis – Discourse Analysis - Introduction to real-life applications of NLP – Introduction to corpora – Corpora Analysis.

Module:2	Morphological Analysis and POS Tagging	7 hours

Sentence Segmentation – Language Specific issues – Text Normalization – Stemming - Inflectional and Derivation Morphology - Morphological Analysis and Generation using finite state transducers - Introduction to POS Tagging, Hidden Markov Models for POS Tagging - Viterbi Decoding for HMM.

Module:3	Syntax Analysis	6 hours

Ambiguities in Syntax Analysis - Issues with Context Free Grammar based parsing- Shallow parsing- Conditional Random Fields (CRF), Dependency Grammar- Dependency Parsing, Neural Network Dependency Parser.

Module:4	Semantic Analysis	7 hours

Semantics - Lexical Semantics- Word Senses - Relations between Senses - Word Sense Disambiguation (WSD) – Word Similarity Analysis using Thesaurus and Distributional methods – Word2vec – fastText word Embedding - Lesk Algorithm – Thematic Roles, Semantic Role labelling - Pragmatics Analysis - Anaphora Resolution.

Module:5	N Gram Modeling & Word Embeddings	7 hours

Vector space representation - The role of language models - Simple N-gram models - Estimating parameters - Evaluating language models - Basic smoothing – Laplace Smoothing – BERT.

Module:6	Applications Of NLP-I	5 hours
Text Categorization: Sentiment Analysis, Named Entity Recognition- Neural Machine Translation (NMT).		
Module:7	**Applications Of NLP-II**	**7 hours**
Text Summarization – Abstractive Summarization – Extractive Summarization – Question Answering Systems (QA) - Information Retrieval based QA – Knowledge Based QA -Question answering Systems using GPT model.		
Module:8	**Recent Trends**	**2 hours**
Guest lectures from Industry and, Research and Development Organizations		
	Total Lecture hours:	**45 hours**

Text Book(s)	
1.	Daniel Jurafsky and James H. Martin "Speech and Language Processing", 3rd edition, Prentice Hall, 2017.
2.	Christopher D. Manning and HinrichSchutze, "Foundations of Statistical Natural Language Processing", The MIT Press – Reprint 2016.
Reference Books	
1.	Hobson Lane, Hannes Hapke, Cole Howard," Natural Language Processing in Action", Manning Publications, 2019.
2.	SowmyaVajjala, Bodhisattwa Majumder, Anuj Gupta, HarshitSurana, Practical Natural Language Processing, O'Reilly Media, Inc. 2020.

Mode of Evaluation: CAT / written assignment / Quiz / FAT / Project			

Recommended by Board of Studies	12-05-2023		
Approved by Academic Council	No. 70	Date	24-06-2023