GUJARAT TECHNOLOGICAL UNIVERSITY

Bachelor of Engineering
Subject Code: 3170723
Semester – VII
Subject Name: Natural Language Processing

Type of course: Elective

Prerequisite: Probability and statistics, Programming and data structures

Rationale: Automated processing of human languages is increasingly becoming important for different types of applications including language translation, surveys, chatbots etc. This subject introduces the fundamentals of natural language processing and its applications in various problem domains.

Teaching and Examination Scheme:

Teaching Scheme			Credits	Examination Marks				Total Marks
L	T	P	C	Theory Marks		Practical Marks		
				ESE (E)	PA (M)	ESE (V)	PA (I)	
3	0	2	4	70	30	30	20	150

Content:

Sr. No.	Content	Total Hrs
1	**Introduction to NLP:** What is NLP? Why NLP is Difficult? History of NLP, Advantages of NLP, Disadvantages of NLP, Components of NLP, Applications of NLP, How to build an NLP pipeline? Phases of NLP, NLP APIs, NLP Libraries	6
2	**Language Modeling and Part of Speech Tagging:** Unigram Language Model, Bigram, Trigram, N-gram, Advanced smoothing for language modeling, Empirical Comparison of Smoothing Techniques, Applications of Language Modeling, Natural Language Generation, Parts of Speech Tagging, Morphology, Named Entity Recognition	12
3	**Words and Word Forms:** Bag of words, skip-gram, Continuous Bag-Of-Words, Embedding representations for words Lexical Semantics, Word Sense Disambiguation, Knowledge Based and Supervised Word Sense Disambiguation	7
4	**Text Analysis, Summarization and Extraction:** Sentiment Mining, Text Classification, Text Summarization, Information Extraction, Named Entity Recognition, Relation Extraction, Question Answering in Multilingual Setting; NLP in Information Retrieval, Cross-Lingual IR	10
5	**Machine Translation:** Need of MT, Problems of Machine Translation, MT Approaches, Direct Machine Translations, Rule-Based Machine Translation, Knowledge Based MT System, Statistical Machine Translation (SMT), Parameter learning in SMT (IBM models) using EM), Encoder-decoder architecture, Neural Machine Translation	10

Suggested Specification table with Marks (Theory):

Distribution of Theory Marks					
R Level	U Level	A Level	N Level	E Level	C Level
7	14	21	14	7	7

Legends: R: Remembrance; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create and above Levels (Revised Bloom's Taxonomy)

Reference Books:

1. Speech and Language Processing: AnIntroduction to Natural Language Processing, Computational Linguistics and Speech Recognition Jurafsky, David, and James H. Martin, PEARSON
2. Foundations of Statistical Natural Language Processing, Manning, Christopher D., and Hinrich Schütze, Cambridge, MA: MIT Press
3. Natural Language Understanding, James Allen. The Benjamin/Cummings Publishing Company Inc..
4. Natural Language Processing with Python – Analyzing Text with the Natural Language ToolkitSteven Bird, Ewan Klein, and Edward Loper.

Course Outcomes:

Sr. No.	CO statement	Marks % weightage
CO-1	Understand comprehend the key concepts of NLP and identify the NLP challenges and issues	14
CO-2	Develop Language Modeling for various text corpora across the different languages	28
CO-3	Illustrate computational methods to understand language phenomena of word sense disambiguation	14
CO-4	Design and develop applications for text or information extraction/summarization/classification.	24
CO-5	Apply different Machine translation techniques for translating a source to target language(s)	20

List of Experiments: Practical work will be based on the above syllabus with minimum 10 experiments to be performed.

List of e-Learning Resources:

1. https://www.kaggle.com/learn/natural-language-processing
2. https://www.javatpoint.com/nlp
3. https://nptel.ac.in/
4. https://www.coursera.org/