GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Bachelor of Engineering

Level: UG

Branch: Computer Engineering

Course / Subject Code : 3164902

CC Course / Subject Name : Natural Language Processing

w. e. f. Academic Year:	2024-2025
Semester:	6th
Category of the Course:	Professional Elective Course

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.

Course Outcome:

After Completion of the Course, Student will able to:

No	Course Outcomes	RBT Level
1	Understand comprehend the key concepts of NLP and identify the NLP challenges and issues	U
2	Develop Language Modeling for various text corpora across the different languages	A
3	Illustrate computational methods to understand language phenomena of word sense disambiguation	U
4	Design and develop applications for text or information extraction/summarization/classification.	A
5	Apply different Machine translation techniques for translating a source to target language(s)	A

**Revised Bloom's Taxonomy (RBT)*

Teaching and Examination Scheme:

Teaching Scheme (in Hours)			Total Credits L+T+ (PR/2)	Assessment Pattern and Marks				Total Marks
				Theory		Tutorial / Practical		
L	T	PR	C	ESE (E)	PA / CA (M)	PA/CA (I)	ESE (V)	
3	0	2	4	70	30	30	20	150

GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Bachelor of Engineering

Level: UG

Branch: Computer Engineering

Course / Subject Code : 3164902

CC Course / Subject Name : Natural Language Processing

Course Content:

Unit No.	Content	No. of Hours	% of Weightage
1.	**Unit 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	12
2.	**Unit 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	28
3.	**Unit 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	8	16
4.	**Unit 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	11	22
5.	**Unit 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.	11	22
	Total	**48**	**100**

Suggested Specification Table with Marks (Theory):

Distribution of Theory Marks (in %)					
R Level	**U Level**	**A Level**	**N Level**	**E Level**	**C Level**
15	**30**	**30**	**15**	**10**	**00**

Where R: Remember; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create (as per Revised Bloom's Taxonomy)

References/Suggested Learning Resources:

(a) 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.

(b) Open source software and website:

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/

* * * * * * *