



GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Master of Engineering

Level: PG

Branch: Information Technology

Subject Code : ME02023011

Course / Subject Name : Natural Language Processing

w. e. f. Academic Year:	A.Y. 2024-25
Semester:	2
Category of the Course:	PCC

Prerequisite:	Data structures& algorithms, finite automata, probability theory and basic programming.
Rationale:	This course aims to provide students with fundamentals and working of NLP and enable student to apply it.

Course Outcome:

After Completion of the Course, Student will able to:

No	Course Outcomes	RBT Level
1	Understand various levels of analysis involved in NLP.	U
2	Understand language generation and discourse analysis	U
3	Able to design, implement, and analyze NLP algorithms	A
4	Explore various application of NLP like Machine Translation, Text Summarization	N

*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
L	T	PR	C	Theory		Tutorial / Practical		
				ESE (E)	PA / CA (M)	PA/CA (I)	ESE (V)	
3	0	2	4	70	30	20	30	150

Course Content:

Unit No.	Content	No. of Hours	% of Weightage
1.	Introduction Introduction to NLP. NLP tasks in syntax, semantics, and pragmatics. Applications such as information extraction, question answering, and machine translation. The problem of ambiguity. The role of machine learning. Brief history of the field.	3	10



GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Master of Engineering

Level: PG

Branch: Information Technology

Subject Code : ME02023011

Course / Subject Name : Natural Language Processing

2.	N-gram Language Models N-grams Models of Syntax - Counting Words - Unsmoothed N-grams. Smoothing- Back-off Deleted Interpolation – Entropy - English Word Classes - Tag sets for English Part of Speech Tagging-Rule Based Part of Speech Tagging - Stochastic Part of Speech Tagging - Transformation-Based Tagging	10	20
3	Context Free Grammars Context Free Grammars for English Syntax- Context-Free Rules and Trees -Understand the network simulation tools. Sentence- Level Constructions–Agreement – Sub Categorization .Parsing – Top-down – Early Parsing -feature Structures – Probabilistic Context-Free Grammars	7	15
4	Semantic Analysis: Representing Meaning-Meaning Structure of Language-First Order Predicate Calculus Representing Linguistically Relevant Concepts -Syntax-Driven Semantic Analysis - Semantic Attachments -Syntax-Driven Analyzer. Robust Analysis - Lexemes and Their Senses - Internal Structure - Word Sense Disambiguation - Information Retrieval	7	15
5.	Language Generation And Discourse Analysis: Discourse - Reference Resolution - Text Coherence -Discourse Structure – Coherence. Dialog and Conversational Agents - Dialog Acts – Interpretation -Conversational Agents. Language Generation– Architecture-Surface Realizations - Discourse Planning .Machine Translation -Transfer Metaphor– Interlingua – Statistical Approaches	7	15
6.	Information Extraction (IE): Named entity recognition and relation extraction. IE using sequence labeling.	3	10
7.	Machine Translation (MT): Basic issues in MT. Statistical translation, word alignment, phrase-based translation, and synchronous grammars.	4	10
8.	Applications of NLP : Sentiment Analysis, Chatbot, Text Summarization, Question/Answering	4	5
TOTAL		45	100

Suggested Specification Table with Marks (Theory):

Distribution of Theory Marks (in %)					
R Level	U Level	A Level	N Level	E Level	C Level
-	40	30	30	--	---

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



GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Master of Engineering

Level: PG

Branch: Information Technology

Subject Code : ME02023011

Course / Subject Name : Natural Language Processing

References/Suggested Learning Resources:

(a) Books:

1. D. Jurafsky and J. Martin “Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition”, Third Edition draft
2. C. Manning and H. Schutze, “Foundations of Statistical Natural Language Processing”, MIT Press, 1999
3. Bharati A., Sangal R., Chaitanya V.. Natural language processing: a Paninian perspective, PHI, 2000
4. Siddiqui T., Tiwary U. S. Natural language processing and Information retrieval, OUP 2008

(b) Open source software and website:

1. NPTEL Course: Natural Language Processing, IIT Kharagpur by Prof. Pawan Goyal

Suggested Course Practical List: (List can be change according to Latest Development)

Sr	Aim
1	Basic Text Processing operation on text document
2	Implement N-gram Language model
3	Write a program to extract features from text
4	Implement word embedding using Word2Vec/Glove/fastText
5	Implement LSA and Topic model
6	Implementation text classification using Naïve Bayes, SVM
7	Implementation of K-means Clustering algorithm on text
8	Implement PoS Tagging on text
9	Implement text processing with neural network
10	Implement text processing with LSTM
11	Implement HMM/CRF on sequence tagging task
12	Develop any one NLP application <ul style="list-style-type: none">• Sentiment Analysis• Chatbot• Text Summarization• Question/Answering
