GUJARAT TECHNOLOGICAL UNIVERSITY
Syllabus for Integrated MSc, 8th Semester
Branch: Information Technology
Subject Name: Natural Language Processing
Subject Code: 1380507

Teaching and Examination Scheme

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

Content:

Sr. No.	Content	Teaching Hours	Module Weightage (%)
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.	5	15%
2.	**Text Processing and Morphology** Character Encoding, Word Segmentation, Sentence Segmentation, Introduction to Corpora, Corpora Analysis Morphology Analysis, Key Techniques used in Morphological Analysis, Applications of Morphological Analysis	7	20%
3.	**Lexical Syntax and Language Modeling** Introduction to word types, POS Tagging, Maximum Entropy Models for POS tagging, Bag of words, Multi- word Expressions, Unigram Language Model, Bigram, Trigram, N-gram, Smoothing Techniques for language modeling, Evaluating language models, Applications of Language Modeling, Named Entity Recognition	10	25%
4.	**Syntax & Semantics** Introduction to phrases, clauses and sentence structure, Shallow Parsing and Chunking, Shallow Parsing with Conditional Random Fields (CRF), Lexical Semantics, Word Sense Disambiguation, WordNet, Thematic Roles, Semantic Role Labelling with CRFs.	8	15%
5.	**Text Analysis, Summarization and Extraction** NL Interfaces, Text Summarization, Sentiment Analysis, Information Extraction, Machine Translation.	10	25%

Reference Books:

1. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition Jura sky, David, and James H. Martin, PEARSON
2. Foundations of Statistical Natural Language Processing, Manning, Christopher D., and

Hinrich Schutzer, 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 Toolkit Steven Bird, Ewan Klein, and Edward Loper

Course Outcome:

After learning the course, the students should be able to:

No.	CO statement
CO-1	Understand comprehend the key concepts of NLP and identify the NLP challenges and issues.
CO-2	Demonstrate understanding of state-of-the-art algorithms and techniques for text-based processing of natural language with respect to morphology.
CO-3	Develop Language Modeling for various text corpora across the different languages.
CO-4	Check the syntactic and semantic correctness of sentences using grammars and labelling.
CO-5	Design and develop applications for text or information extraction / summarization /classification.