



# GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Post Graduate Diploma In Data Science

Level: PG

Branch: Data Science

Subject Code: DS02080071

Subject Name: Natural Language Processing

w. e. f. Academic Year:	2024-25
Semester:	2
Category of the Course:	Professional Elective - I

<b>Prerequisite:</b>	Nil
<b>Rationale:</b>	Nil

## Objectives

- To make the students aware about processing of unstructured data.
- To design the computer programs that consider and process the textual data as human-like comprehension.
- To design the computer system that able to understand and draw inferences from textual data.

## Course Outcomes:

Sr. No.	CO statement
1	After completion of subject, students would be able to:
2	Understand fundamental mathematical models and algorithms in the field of NLP.
3	Perform lexical, syntax and semantic analysis of textual data.
4	Apply these concepts in real world problems like sentiment analysis, speech recognition, automated question answering, text classification etc.

## Teaching and Examination Scheme:

Teaching / Learning Scheme (in Hours per semester)			Total Credits	Assessment Pattern and Marks				Total Marks
L	T	PR		Theory		Tutorial / Practical		
				ESE (E)	PA / CA (M)	PA/CA (I)	ESE (V)	
4	0	2	5	70	30	0	50	150

## Content:

Sr. No.	Content	Total Hrs
1.	Introduction What is Natural Language Processing, Ambiguity and uncertainty in language. The Turing test, Models and Algorithms	2
2.	Regular Expressions, Automata and Finite-State Transducers Basic Regular Expression Patterns, Disjunction, Grouping, and Precedence, Examples, Advanced Operators, Regular Expression Substitution, Using FSA	10



# GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Post Graduate Diploma In Data Science

Level: PG

Branch: Data Science

Subject Code: DS02080071

Subject Name: Natural Language Processing

	to Recognize Sheeptalk, Formal Languages, Nondeterministic FSAs, Using an NFSA to accept strings, Recognition as Search, Relating Deterministic and Non-deterministic Automata, Regular Languages and FSAs, Survey of (Mostly) English Morphology, Inflectional Morphology, Derivational Morphology, The Lexicon and Morphotactics, Morphological Parsing with Finite-State Transducers, Orthographic Rules and Finite-State Transducers, Human Morphological Processing	
3.	N-gram Language Models Words, Corpora, Text Normalization, Minimum Edit Distance, N-Grams, Evaluating Language Models, Generalization and Zeros, Smoothing, Kneser-Ney Smoothing	4
4.	Word Classes and Part-of-Speech Tagging English Word Classes, Tagsets for English, Part of Speech Tagging, Rule-based Part-of-speech Tagging, Stochastic Part-of-speech Tagging, The Actual Algorithm for HMM tagging, Transformation-Based Tagging, How TBL rules are applied, How TBL Rules are Learned, Multiple tags and multiple words, Unknown words, Class-based N-grams	6
5.	Context-Free Grammars for English and Parsing Constituency, Context-Free Rules and Trees, Sentence-Level Constructions, The Noun Phrase, The Verb Phrase and Subcategorization, Auxiliaries, Spoken Language Syntax, Grammar Equivalence & Normal Form, Finite State & Context-Free Grammars, Grammars & Human Processing, Parsing as Search, Top-Down and Bottom-Up Parsing with comparison, A Basic Top-down Parser, Adding Bottom-up Filtering, Problems with the Basic Top-down Parser, Left Recursion, Ambiguity	10
6.	Semantics and Pragmatics Syntax-Driven Semantic Analysis, Attachments for a Fragment of English, Idioms and Compositionality, Robust Semantic Analysis, WordNet: A Database of Lexical Relations, Reference Resolution, Text Coherence, Discourse Structure, Psycholinguistic Studies of Reference and Coherence,	8
<b>TOTAL</b>		

## Reference Books:

1	Speech and Language Processing by Daniel Jurafsky and James H. Martin Second edition, Prentice Hall
2	Natural Language Understanding by Allen James Second edition, Benjamin/Cumming
3	Statistical Language Learning by Charniack Eugene MIT Press
4	Foundations of Statistical Natural Language Processing by Manning Christopher and Heinrich Schutze MIT Press

\*\*\*\*\*