

Course Objective: 1. To familiarize with basic text pre-processing steps and pattern matching using regular expressions
2. To design grammar based text parsing systems.
3. To understand, design and implement machine learning based text classification and text generation systems.

S. NO.	Course Outcomes (CO)
CO1	Ability to understand and implement text pre-processing techniques and pattern matching using regular expressions.

CO2	Ability to implement Bag-of-Words models for feature extraction and machine learning for text classification.
CO3	Ability to analyze the use of word senses and word embeddings with sequence learning using LSTM or transformer.
CO4	Ability to execute syntactic parsing of sentences for a given grammar and probabilistic inferencing.
CO5	Ability to design and implement text generation models using Language modelling and encoder-decoder models.
CO6	Ability to describe real-world applications of Natural Language Processing and Natural Language Generation.

S. NO.	Contents	Contact Hours
UNIT 1	Introduction: The study of Language, Introduction to NLP and various terms related to NLP- morphology, syntax, semantics, pragmatics, discourse, ambiguity. Regular Expression, Finite State Automata.	6
UNIT 2	Text representation and classification: Pre-processing: Tokenization, Lemmatization, Stemming. Frequency Based Methods: Bag-of-Words features, 1-hot encoding, TF, TF-IDF, Machine learning classifiers. Sequence Based Methods: Word Embeddings, LSTM/Transformer.	8
UNIT 3	Lexical Semantics: Word Senses, WordNet, SentiWordNet, Synsets, Hypernyms, Hyponyms, Meronyms, Holonyms, Word Sense Disambiguation, Word Similarity, Semantic Role Labelling.	8
UNIT 4	Natural language generation: Probabilistic Context-Free Grammars, Syntactic Parsing, Part-of-speech-tagging, Probabilistic Language Processing, N-gram language modelling for text generation. Encoder-decoder LSTM/Transformer models for sequence-to-sequence learning, Attention mechanism.	12
UNIT 5	Advanced techniques: Large Language Models, Machine Translation, Man-Machine Interfaces, Natural language Question-Answering Systems, Text summarization models, Information Retrieval.	8
	TOTAL	42

REFERENCES

S.No.	Name of Books/Authors/Publishers	Year of Publication / Reprint
1	D. Jurafsky, J. H. Martin, Speech and Language Processing, 2/e, Pearson Education.	2013
2	James Allen, Natural Language Understanding, 2/e, Pearson Education.	2003