



Gyanmanjari
Innovative University

Course Syllabus
Gyanmanjari Institute of Technology
Semester-6 (B. Tech.)

Subject: Natural Language Processing-BETCE15402

Type of course: Professional Core

Prerequisite: Prerequisites include programming (Python), data structures, algorithms, probability, statistics, and basic machine learning concepts, with preferred understanding of linguistics—grammar, syntax, and semantics for NLP applications.

Rationale:

This course introduces the foundations and applications of Natural Language Processing (NLP), enabling students to analyze and understand human language using computational techniques. Covering preprocessing, language modeling, semantics, and deep learning, it prepares students to build intelligent systems like chatbots and translation tools, aligning with modern AI industry and research trends.

Teaching and Examination Scheme:

Teaching Scheme			Credits	Examination Marks					Total Marks
CI	T	P		C	Theory Marks		Practical Marks		
			ESE		MSE	V	P	ALA	
4	0	2	5	60	30	10	20	30	150

Legends: CI-ClassRoom Instructions; T – Tutorial; P - Practical; C – Credit; ESE - End Semester Examination; MSE- Mid Semester Examination; V – Viva; CA - Continuous Assessment; ALA- Active Learning Activities.



Course Content:

Sr. No	Course Content	Hrs.	% Weightage
1	Introduction to NLP & Text Processing: Overview of Natural Language Processing, Scope and Challenges of NLP, Evolution of NLP and its milestones; Pros and Cons of NLP systems; Core building blocks of NLP, Major application domains of NLP; Steps in developing an NLP pipeline, Processing stages in NLP systems, Introduction to commonly used NLP Models (Neural network models (RNN, LSTM, GRU). and open-source libraries.	12	20%
2	Language Models & Syntax Parsing: N-gram language models (Unigram, Bigram, Trigram), smoothing techniques, Basics of natural language generation; Introduction to Part-of-Speech tagging and its applications, Part-of-Speech tagging methods (rule-based, stochastic, transformation-based).	12	20%
3	Word Representations and Semantics: Bag-of-Words model, Vector-space representations of words; Word embedding techniques (Skip-gram, CBOW); Distributed semantic representations; Basics of lexical semantics; Word Sense Disambiguation approaches and applications.	12	20%
4	Text Processing, Analysis and Extraction: Core concepts of Information Extraction, Text mining fundamentals, Sentiment analysis and opinion mining; Document classification techniques; Text summarization methods; Named Entity Recognition and Relation Extraction.	12	20%
5	Applications & Deployment: Question answering systems, text summarization, chatbots, information retrieval. Basics of Statistical Machine Translation; Brief about Neural Machine Translation and modern MT architectures.	12	20%



Continuous Assessment:

Sr. No	Active Learning Activities	Marks
1	Exploring N-Gram Models with Smoothing: Students will implement word-level N-Gram models in Python individually, apply smoothing techniques (Laplace/Back-off), and compare the performance on different datasets. They will prepare documentation of their observations in pdf format and submit the final report on the GMIU Web Portal.	10
2	TF-IDF Implementation: Students have to individually implement a Bag of Words and TF-IDF model on a sample text dataset. Submit code and short analysis in pdf format and then students have to upload it on GMIU Portal.	10
3	Sentiment Analyzer / Chatbot Prototype: Students will develop a mini-project individually implementing either a sentiment analysis system or a rule-based chatbot using NLP libraries (NLTK, SpaCy, HuggingFace). They will demonstrate the prototype in class and upload the project documentation in pdf format on the GMIU Web Portal.	10
Total		30

Suggested Specification table with Marks (Theory):60

Distribution of Theory Marks (Revised Bloom's Taxonomy)						
Level	Remembrance (R)	Understanding (U)	Application (A)	Analyze (N)	Evaluate (E)	Create (C)
Weightage %	25%	25%	15%	15%	10%	10%

Note: This specification table shall be treated as a general guideline for student and teacher. The actual distribution of marks in the question paper may vary slightly from above table.



Course Outcome:

After learning the course the students should be able to:	
CO1	Explain the fundamentals of Natural Language Processing, its challenges, ambiguity issues, and applications such as machine translation, question answering, and information extraction.
CO2	Apply text preprocessing techniques and implement N-Gram language models with smoothing and part-of-speech tagging.
CO3	Analyze and represent textual data using Bag-of-Words, TF-IDF, and Word Embedding models for various NLP tasks.
CO4	Implement text analysis techniques such as sentiment analysis, text classification, and named entity recognition using standard NLP tools.
CO5	Develop simple NLP applications using common libraries and comprehend the core ideas behind modern NLP systems, including chatbots and machine translation models.

List of Practical

Sr. No	Description	Unit No	Hrs.
1	Install NLTK and Perform Basic Text Operations Tokenize a paragraph into sentences and words. Remove stop words and punctuation from the text.	1	02
2	Count the Number of Nouns, Verbs, Adjectives, and Adverbs in a Given Paragraph Using POS Tagging	2	04
3	Read a text file (e.g., a short story). Create a frequency distribution of words and display the 10 most common words.	2	02
4	From a small text corpus, generate and print all possible bigrams. Calculate the frequency of 5 specific bigrams.	2	02
5	Calculate Word Frequency (Term Counts) for Each Sentence and Print the Most Common Words	3	02
6	Given 3 simple sentences, create a BoW representation using scikitlearn's CountVectorizer and print the vocabulary and word counts.	3	02
7	Use TextBlob to analyze the sentiment of 5 user-input sentences. Print whether each sentence is Positive, Negative, or Neutral.	4	04
8	Extract Nouns, Verbs, and Adjectives from a Paragraph Using spaCy and Display the Counts	4	04
9	Build a Simple Rule-Based Question Answering System Using Python Dictionaries.	5	04
10	Implement a Simple Keyword-Based Search Engine That Finds Relevant Sentences From a Document.	5	04
Total			30



Instructional Method:

The course delivery method will depend upon the requirement of content and need of students. The teacher in addition to conventional teaching method by black board, may also use any of tools such as demonstration, role play, Quiz, brainstorming, MOOCs etc.

From the content 10% topics are suggested for flipped mode instruction.

Students will use supplementary resources such as online videos, NPTEL/SWAYAM videos, e-courses, Virtual Laboratory.

The internal evaluation will be done on the basis of Active Learning Assignment.

Practical/Viva examination will be conducted at the end of semester for evaluation of performance of students in laboratory.

Reference Books:

- [1] Daniel Jurafsky, James H. Martin – Speech and Language Processing – Pearson, 3rd Edition, 2023.
- [2] Christopher D. Manning, Hinrich Schütze – Foundations of Statistical Natural Language Processing – MIT Press, 1999.
- [3] Steven Bird, Ewan Klein, Edward Loper – Natural Language Processing with Python – O'Reilly Media, 2009.
- [4] Yoav Goldberg – Neural Network Methods for Natural Language Processing – Morgan & Claypool Publishers, 2017.
- [5] Nitin Indurkha, Fred J. Damerau – Handbook of Natural Language Processing – Chapman and Hall/CRC, 2nd Edition, 2010.

