

Branch Name:	Bachelor of Technology (Computer Engineering / Information Technology)
Program Code:	ET103/ET105
Course Title:	Elective -IV Practical: Natural Language Processing Practical
Course Code:	2ET1030704P
Pre-requisite Course:	Basics of Artificial Intelligence & Machine Learning

Course Objectives: The course objectives for Natural Language Processing are designed to bridge the gap between human communication and computational logic, Linguistic Proficiency, Grammatical & Semantic Analysis, Mathematical Modelling, Deep Learning Mastery, Practical Application & Evaluation.

Teaching and Examination Scheme:

Credit	Teaching Scheme Hours per week				Evaluation Scheme (Marks)								
					Theory (Marks)				Practical (Marks)				Total (Marks)
	L	T	P	Total	SEE		CCA		SEE		CCA		
					Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	
1	-	-	2	2	-	-	-	-	10	25	10	25	50

List of Experiments:

S.N.	Practical Exercise
1	Text Preprocessing Pipeline: Write a Python script to perform the following on a raw text corpus: Sentence and Word Tokenization, Removal of stop-words and special characters, Comparison between Porter Stemmer and WordNet Lemmatizer.
2	Regular Expressions for Information Extraction: Create a system using RegEx to extract specific patterns from a document, such as: Email addresses, Phone numbers, and Dates, Finding all hashtags or mentions in a Twitter dataset.
3	N-gram Language Modeling: Implement a Bigram and Trigram model to: Calculate the probability of a given sentence, Generate a random sequence of 10 words based on the frequency distribution of a provided text file.
4	Parts-of-Speech (POS) Tagging: Use a pre-trained POS tagger to label a paragraph. Compare the accuracy of a Rule-based tagger vs. a Stochastic tagger (HMM).
5	Named Entity Recognition (NER): Extract entities (Persons, Organizations, and Locations) from a news article. Visualize the entities using spaCy's display visualizer.
6	Text Classification (Sentiment Analysis): Build a movie review classifier (Positive/Negative) using: Feature Extraction: TF-IDF, Classifier: Naive Bayes or Support Vector Machine (SVM) using Dataset: IMDB Movie Reviews.
7	Word Embedding: Use Gensim to load a pre-trained Word2Vec or GloVe model. Find the top 5 words similar to "Technology", Perform word arithmetic: King - Man + Woman = ?
8	Sequence Modeling with RNN/LSTM: Build a simple Next Word Predictor or a Spam Classifier using an LSTM network in PyTorch or Keras.
9	Machine Translation with Encoder-Decoder: Implement a basic English-to-French (or any local language) translator using a Sequence-to-Sequence model with an Attention Mechanism.
10	Transformer Fine-Tuning: Use the Hugging Face library to: Load a pre-trained BERT model, Fine-tune it for a specific task like Question Answering or Document Summarization.

Reference Books:

1. Speech and Language Processing by Dan Jurafsky and James H., Pearson
2. Foundations of Statistical Natural Language Processing by Christopher Manning & Hinrich Schütze, MIT Press
3. Natural Language Understanding by James Allen, Pearson Education
4. Natural Language Processing with Python by Steven Bird, Ewan Klein & Edward Loper O'Reilly Media

List of Open-Source Software/learning website:

1. <https://stanfordnlp.github.io/CoreNLP/>
2. <https://www.nltk.org/>
3. <https://www.coursera.org/specializations/natural-language-processing>
4. <https://aclanthology.org/>
5. <https://archive.ics.uci.edu/>

Course Outcomes (CO):

Course Outcomes	Description	Bloom's Taxonomy Level
CO1	Understand the linguistic challenges, ambiguity, and the standard processing pipeline.	Remembering (1) Understanding (2)
CO2	Execute tokenization, stemming, lemmatization, and use FSTs for word-level analysis.	Applying (3)
CO3	Evaluate sentences using CFGs, dependency parsing, and resolve word sense ambiguity.	Analyze (4)
CO4	Implement TF-IDF, Word2Vec, and GloVe models to transform text into numerical features.	Applying (3) Create (6)
CO5	Build and train RNN, LSTM, GRU, and BERT architectures for language modeling and translation.	Evaluate (5) Create (6)

Mapping of CO with PO and PSO

Course Outcomes	Program Outcomes (POs)												Program Specific Outcomes(PSOs)			
	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PSO 1	PSO 2	PSO 3	PSO 4
CO1	H	M	-	-	-	-	-	-	-	-	-	L	M	-	-	-
CO2	H	M	M	-	M	-	-	-	-	-	-	-	H	M	-	-
CO3	H	H	M	M	-	-	-	-	-	-	-	-	M	M	-	-
CO4	H	H	H	M	H	-	-	-	-	-	-	M	H	H	-	-
CO5	M	H	H	H	H	-	-	-	M	-	-	M	H	H	-	-

H: High, M: Medium, L: Low