Batch: C1	Roll No.: 16010122323
Experiment / assignment / tutorial No. 9	

Title: Applications of NLP

Objective: Self-Learning Experiment in NLP to introduce concepts like word analysis, word generation and building POS Tagger

Expected Outcome of Experiment:

Course Outcom	After successful completion of the course students should be able to
CO4	Analyse applications of AI and understand planning & learning processes in advanced AI applications

Books/ Journals/ Websites referred:

1. **"Artificial Intelligence: a Modern Approach" by Russel and Norving, Pearson education Publications**
2. **"Artificial Intelligence" By Rich and knight, Tata Mcgraw Hill Publications**

Pre Lab/ Prior Concepts: Natural Language Processing

Introduction: Natural Language Processing (NLP) is a dynamic subfield of artificial intelligence (AI) and linguistics that empowers machines to comprehend, interpret, generate, and respond to human language in a meaningful way. It bridges the gap between computational understanding and natural human communication, enabling computers to process and derive insights from large volumes of unstructured textual or spoken data.

NLP combines techniques from computer science, linguistics, machine learning, and deep learning to model and analyze language phenomena. It handles both syntax (structure) and semantics (meaning), enabling systems to go beyond simple keyword matching toward contextual and intent-aware understanding.

Key Applications of NLP

NLP has become an integral part of numerous real-world applications, such as:

- **Text Classification**: Categorizing text into predefined labels (e.g., spam detection, fake news detection).

- **Text Summarization**: Generating concise versions of longer documents while preserving key information.
- **Sentiment Analysis**: Determining the emotional tone behind text, commonly used in product reviews and social media monitoring.
- **Machine Translation**: Translating text or speech from one language to another (e.g., Google Translate).
- **Speech Recognition**: Converting spoken language into written text, essential for voice assistants like Siri or Alexa.
- **Named Entity Recognition (NER)**: Identifying and classifying entities (e.g., names, dates, locations) within text.
- **Question Answering**: Building intelligent systems that can understand and respond to natural language queries.
- **Chatbots and Conversational Agents**: Enabling natural conversations between humans and machines.
- **Topic Modeling and Information Extraction**: Automatically detecting key themes, topics, or facts from large text corpora.

Recent Advances in NLP

In recent years, NLP has experienced a major leap forward due to the rise of **transformer-based models** like BERT, RoBERTa, GPT, and T5. These models utilize attention mechanisms to capture context more effectively, enabling state-of-the-art performance in tasks like fake news detection, abstractive summarization, and even creative text generation.

Pre-trained language models have made it possible to fine-tune systems for specific tasks with limited labeled data, accelerating the development of NLP applications across various domains such as healthcare, legal tech, journalism, and finance.

Application 1

Theory: Discuss/Add the following details about the Project/ Application chosen.

Title:
Sentiment Analysis of Product Reviews

Introduction:

Sentiment analysis is one of the most common applications of NLP, used to determine whether a piece of text expresses positive, negative, or neutral emotions. It is widely used in customer feedback systems, brand monitoring, and product reviews.

Methodology (Includes Data Description, Models used , Tools and Libraries ,etc) :

Data Description:
A dataset of Amazon product reviews including user comments and star ratings.

Models Used:
Logistic Regression, Naive Bayes, and LSTM (Long Short-Term Memory) neural networks for deep learning.

Tools and Libraries:
Python, NLTK, Scikit-learn, TensorFlow/Keras, and pandas for preprocessing.Steps:

1. Clean and tokenize review text.
2. Use TF-IDF vectorization.
3. Train sentiment classification model.
4. Evaluate accuracy and F1-score.

Discussion:

The Logistic Regression model performed well on short reviews, but LSTM provided better accuracy on long and complex sentences. Feature engineering (like using bigrams or stop-word filtering) improved performance significantly. Word embeddings such as GloVe were also tested for better semantic understanding.

Conclusion:
Sentiment analysis helps companies understand user opinions at scale. The NLP pipeline from text preprocessing to deep learning classification can be tuned to achieve reliable performance for real-world applications.

Application 2

Theory: Discuss/Add the following details about the Project/ Application chosen.

Title:
Named Entity Recognition (NER) for News Articles

Introduction:

Named Entity Recognition is a task in NLP that identifies and categorizes entities like names of people, organizations, locations, dates, etc., in text. It is critical for applications like automated news summarization, search engines, and knowledge graph construction.

Methodology (Includes Data Description, Models used , Tools and Libraries ,etc):

Data Description:
CoNLL-2003 dataset and a custom news articles corpus from online sources.

Models Used:
Conditional Random Fields (CRF), spaCy's pretrained models, and BERT-based transformers.

Tools and Libraries:
spaCy, HuggingFace Transformers, scikit-learn, pandas, Python.

Steps:

1. Load and preprocess data (tokenization, POS tagging).
2. Annotate text with entity labels (PER, ORG, LOC, MISC).
3. Train CRF and fine-tune BERT on entity tags.
4. Evaluate with precision, recall, and F1-score.

Discussion:

Pretrained transformer models like BERT showed a major improvement in identifying complex and nested entities. CRF models were faster but less accurate on unseen data. Transfer learning using HuggingFace allowed quick adaptation to domain-specific text like financial or legal news.

Conclusion:

NER enables extraction of structured information from unstructured text. Modern deep learning-based models significantly outperform traditional approaches, especially on large and varied datasets.

Reference Link 1: https://www.nltk.org/book/ch06.html

Reference Link 2 : https://huggingface.co/transformers/task_summary.html#named-entity-recognition