

PRACTICAL 1 : NER

Aim:

The aim of this assignment is to explore the use of Named Entity Recognition (NER) as an information extraction technique in natural language processing. It focuses on identifying and classifying key entities in text data to enhance data analysis and understanding.

Objectives:

1. To define Named Entity Recognition and its significance in natural language processing.
2. To evaluate various NER algorithms and models available for entity extraction.
3. To implement a practical NER application using a suitable programming language and library.

Theory:

Named Entity Recognition (NER) is a crucial technique in information extraction that identifies and classifies named entities in unstructured text into predefined categories, such as persons, organizations, locations, dates, and more. By leveraging NER, researchers and practitioners can extract meaningful insights from large datasets, enabling more efficient data processing and analysis.

PRACTICAL 2 : SENTIMENT ANALYSIS

Aim:

The aim of this assignment is to implement a sentiment analysis technique for classifying textual data into positive, negative, or neutral sentiment categories. This classification will help in understanding public opinion and emotional tone in various textual datasets.

Objectives:

1. To define sentiment analysis and its relevance in understanding textual emotions and opinions.
2. To explore various sentiment analysis algorithms and libraries used for text classification.
3. To develop a sentiment analysis model and evaluate its performance on a sample dataset.

Theory:

Sentiment analysis is a computational technique used to determine the emotional tone behind a body of text. By classifying text into positive, negative, or neutral sentiments, it provides valuable insights into opinions, attitudes, and emotions expressed in social media, reviews, and other textual data sources. Various methods, including machine learning and deep learning techniques, are employed to build effective sentiment analysis models.

PRACTICAL 3 : TEXT SUMMARIZATION

Aim:

The aim of this assignment is to explore and implement Natural Language Processing (NLP) techniques for text summarization. This process will enable the extraction of key information from larger texts, facilitating easier comprehension and efficient information retrieval.

Objectives:

1. To understand different approaches to text summarization, including extractive and abstractive methods.
2. To investigate various NLP libraries and tools used for text summarization tasks.
3. To develop and evaluate a text summarization model using a sample dataset, assessing its effectiveness in retaining essential information.

Theory:

Text summarization is a Natural Language Processing (NLP) task that aims to condense a lengthy text into a shorter version while preserving its essential information. It can be categorized into extractive summarization, which selects key sentences from the text, and abstractive summarization, which generates new sentences to convey the main ideas. By employing various NLP techniques, summarization aids in quickly grasping the content and relevance of documents in an increasingly information-rich environment.

PRACTICAL 4 : MACHINE TRANSLATION

Aim:

The aim of this assignment is to implement a simple machine translation system that can convert text from one language to another. This will demonstrate the fundamental principles of Natural Language Processing (NLP) in translating linguistic content.

Objectives:

1. To study the basic concepts of machine translation, including rule-based and statistical methods.
2. To explore various NLP libraries and frameworks that facilitate machine translation tasks.
3. To develop a prototype translation model and evaluate its accuracy and efficiency using a sample dataset.

Theory:

Machine translation is a subfield of Natural Language Processing (NLP) focused on automatically converting text from one language to another. Traditional approaches include rule-based methods, which rely on linguistic rules, and statistical methods that leverage bilingual text corpora for translation. Recent advancements involve neural machine translation, utilizing deep learning techniques to improve translation quality and fluency. By implementing a simple translation model, one can understand the complexities involved in aligning languages and the challenges of preserving meaning across linguistic boundaries.

PRACTICAL 5 : ASPECT MINING AND TOPIC MODELING

Aim

The aim of this project is to implement aspect mining and topic modeling techniques for extracting meaningful insights from textual data. By identifying key aspects and underlying topics, we can better understand the sentiments and themes present in the text.

Objectives

1. To utilize Natural Language Processing (NLP) techniques for extracting aspects from text data.
2. To implement topic modeling using algorithms like LDA to discover latent topics in the dataset.
3. To evaluate the effectiveness of the aspect mining and topic modeling techniques through qualitative analysis of results.

Theory

Aspect mining focuses on identifying specific features or components within a given text that are of interest, such as product attributes or sentiment-related topics. Topic modeling is a statistical method that analyzes text to uncover hidden thematic structures, helping to group similar content and facilitate better information retrieval and understanding.

PRACTICAL 6 : ADVANCED TOKENIZATION TECHNIQUES

Aim

The aim of this project is to explore and implement advanced tokenization techniques using Python for Natural Language Processing (NLP). By applying various tokenization methods, we seek to enhance the accuracy of text analysis and improve the performance of NLP models.

Objectives

1. To implement and compare basic word tokenization and advanced methods like subword and byte pair encoding.
2. To evaluate the effectiveness of different tokenization techniques on diverse text samples.
3. To demonstrate the practical applications of tokenization in tasks such as sentiment analysis and text summarization.

Theory

Tokenization is a crucial step in NLP that involves splitting text into smaller units, such as words or subwords, for analysis. Advanced tokenization techniques, like subword tokenization, improve model performance by addressing the limitations of traditional methods, such as handling out-of-vocabulary words and improving representation for morphological variations.