Vivekanand Education Society's
Institute Of Technology
Approved by AICTE & Affiliation to University of Mumbai

Artificial Intelligence and Data Science Department
Natural Language Processing / Odd Sem 2023-24 / Experiment 8

Name : Yash Sarang	Class/Roll No. : D16AD / 47	Grade :

Title of Experiment :
Named Entity Recognition in Natural Language Processing (NLP)

Problem Statement :
　　　The problem is to develop a Named Entity Recognizer (NER) that can identify and categorize named entities such as persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc., in a given text. The objective is to build an accurate and efficient NER model to aid in various NLP applications.

Description / Theory :
　　　Named Entity Recognition (NER) is a subtask of information extraction that aims to identify and classify named entities into predefined categories such as person names, organizations, locations, expressions of times, etc. NER is a crucial component in many NLP applications, including question-answering systems, sentiment analysis, and more. Common approaches for NER include rule-based systems, statistical models, and deep learning-based models.

Flowchart :
1. Input the text data.
2. Preprocess the text (tokenization, sentence splitting, etc.).
3. Apply a NER algorithm (rule-based, statistical, deep learning) to identify named entities.
4. Categorize the named entities into predefined categories (person, organization, location, etc.).
5. Output the identified named entities and their categories.

Vivekanand Education Society's
Institute Of Technology
Approved by AICTE & Affiliation to University of Mumbai

Artificial Intelligence and Data Science Department
Natural Language Processing / Odd Sem 2023-24 / Experiment 8

Program:

```python
import spacy


# Load the pre-trained English NER model
nlp = spacy.load("en_core_web_sm")


def named_entity_recognition(text):
    # Process the input text using the spaCy NLP pipeline
    doc = nlp(text)


    # Iterate over the named entities in the processed document
    for ent in doc.ents:
        print(f"Entity: {ent.text}, Type: {ent.label_}")


# Sample text for named entity recognition
sample_text = "Barack Obama was born in Hawaii."


# Perform named entity recognition on the sample text
print("Named Entity Recognition results:")
named_entity_recognition(sample_text)
```

Output:
Named Entity Recognition results:
Entity: Barack Obama, Type: PERSON
Entity: Hawaii, Type: GPE

Results and Discussions :

The results will showcase the identified named entities along with their respective categories using the NER model. The discussion will evaluate the accuracy, recall, precision, and F1-score of the NER model and compare it with other state-of-the-art

Vivekanand Education Society's
Institute Of Technology

Approved by AICTE & Affiliation to University of Mumbai

Artificial Intelligence and Data Science Department
Natural Language Processing / Odd Sem 2023-24 / Experiment 8

approaches. Additionally, the discussion will highlight the strengths and weaknesses of the chosen NER approach.

Conclusion:

Named Entity Recognition (NER) is a critical task in NLP, allowing for the identification and categorization of important named entities in text data. Various algorithms and approaches can be employed to achieve accurate and efficient NER. The successful implementation of an NER model significantly contributes to enhancing the performance of downstream NLP applications and facilitates better information extraction and understanding of text.

******