Vidyavardhini's College of Engineering and Technology

Department of Artificial Intelligence & Data Science

AY:2024-25

Class:	**BE**	**Semester:**	**VII**
Course Code:	**CSDOL7011**	**Course Name:**	**Natural Language Processing**

Name of Student:	**Parth Raut**
Roll No.:	**40**
Experiment No.:	**7**
Title of the Experiment:	**Named Entity Recognition Implementation**
Date of Performance:	
Date of Submission:	

Evaluation

Performance Indicator	Max. Marks	Marks Obtained
Performance	5	
Understanding	5	
Journal work and timely submission	10	
Total	20	

Performance Indicator	Exceed Expectations (EE)	Meet Expectations (ME)	Below Expectations (BE)
Performance	4-5	2-3	1
Understanding	4-5	2-3	1
Journal work and timely submission	8-10	5-8	1-4

Checked by

Name of Faculty :

Signature :

Date :

Experiment 7

Aim: Implement Named Entity Recognizer for the given text input.

Objective: Understand the importance of NER in NLP and Implement NER.

Theory:

The named entity recognition (NER) is one of the most data preprocessing task. It involves the identification of key information in the text and classification into a set of predefined categories. An entity is basically the thing that is consistently talked about or refer to in the text.

NER is the form of NLP.

At its core, NLP is just a two-step process, below are the two steps that are involved:

- Detecting the entities from the text
- Classifying them into different categories

Some of the categories that are the most important architecture in NER such that:

- Person
- Organization
- Place/ location

Other common tasks include classifying of the following:

- date/time.
- expression
- Numeral measurement (money, percent, weight, etc)
- E-mail address

Ambiguity in NE

For a person, the category definition is intuitively quite clear, but for computers, there is some ambiguity in classification. Let's look at some ambiguous example:

England (Organisation) won the 2019 world cup vs The 2019 world cup happened in England(Location).

Washington(Location) is the capital of the US vs The first president of the US was Washington(Person).

Code:

In [1]:
```
import spacy
```

Load the spaCy language model

In [3]:
```
nlp = spacy.load("en_core_web_sm")
```

Sample text input

In [4]:
```
text = "Apple Inc. is a company based in Cupertino, California. John works for Google in Mountain View."
```

Process the text using spaCy

In [5]:
```
import spacy

# Load the spaCy language model
nlp = spacy.load("en_core_web_sm")

# Sample text input
text = "Apple Inc. is a company based in Cupertino, California. John works for Google in Mountain View."

# Process the text using spaCy
doc = nlp(text)

# Initialize variables to store named entities
named_entities = []

# Define a function to extract named entities
def extract_named_entities(doc):
    entities = []
    current_entity = None

    for token in doc:
        if token.ent_type_:
            if current_entity and token.ent_type_ == current_entity[1]:
                current_entity = (current_entity[0] + " " + token.text, token.ent_type_)
            else:
                if current_entity:
                    entities.append(current_entity)
                current_entity = (token.text, token.ent_type_)
        else:
            if current_entity:
                entities.append(current_entity)
            current_entity = None

    if current_entity:
        entities.append(current_entity)

    return entities

# Extract named entities
named_entities = extract_named_entities(doc)

# Print the named entities
for entity, label in named_entities:
    print(f"Entity: {entity}, Label: {label}")
```

```
Entity: Apple Inc., Label: ORG
Entity: Cupertino, Label: GPE
Entity: California, Label: GPE
Entity: John, Label: PERSON
Entity: Google, Label: ORG
Entity: Mountain View, Label: GPE
```

Conclusion:

Implementation of named entity recognition using spaCy successfully extracts various entities from the given text. Here's a breakdown of the results and comments on the identified named entities:

Extracted Named Entities

1. Apple Inc. - Label: ORG
    - Correctly identified as an organization. No issues here.
2. Cupertino - Label: GPE
    - Correctly identified as a geographical entity (GPE). Cupertino is a city in California.
3. California - Label: GPE
    - Correctly identified as a geographical entity (GPE). California is a state in the U.S.
4. John - Label: PERSON
    - Correctly identified as a person. This is accurate based on the context provided.
5. Google - Label: ORG
    - Correctly identified as an organization. This is also accurate.
6. Mountain View - Label: GPE
    - Correctly identified as a geographical entity (GPE). Mountain View is another city in California.

Comments on Results

- The implementation appears to function correctly in recognizing all relevant named entities in the provided text. The entities have been classified accurately according to spaCy's entity types.
- There are no misclassifications or omissions in this instance. All key organizations, persons, and geographical locations mentioned in the text are correctly labeled.