Abhishek Jani **IT Department**

Roll No - 50

Experiment No.5
Perform Natural language Entity Extraction from medical reports
Date of Performance:
Date of Submission:

Aim: Perform Natural language Entity Extraction from medical reports

Objective: The objective of this experiment is to named entity recognition (NER) using SpaCy and related libraries.

Theory:

NER stands for Named Entity Recognition, which is a subtask of Natural Language Processing (NLP). It involves identifying and classifying named entities (such as names of people, organizations, locations, dates, and more) within a text. The goal of NER is to extract structured information from unstructured text data and to recognize specific entities mentioned in the text.

For example, given the sentence: "Apple Inc. was founded by Steve Jobs in Cupertino on April 1, 1976," a named entity recognition system would identify and categorize "Apple Inc." as an organization, "Steve Jobs" as a person, "Cupertino" as a location, and "April 1, 1976" as a date.

NER has various applications in NLP, including information retrieval, question answering, sentiment analysis, text summarization, and more. It plays a crucial role in understanding the context and semantics of a text by identifying and categorizing entities, which can

help in extracting meaningful insights from large amounts of text data.
spaCy:
spaCy is a fast and efficient NLP library that provides pre-trained models for NER. It's known for its speed and accuracy.

```
python
import spacy

nlp = spacy.load("en_core_web_sm")
doc = nlp("Barack Obama was born in Hawaii.")

for ent in doc.ents:
    print(ent.text, ent.label_)
```

Using Named Entity Recognition (NER) in Natural Language Processing (NLP) offers several advantages that enhance the understanding and processing of text data:

Information Extraction: NER allows you to extract structured information from unstructured text. This is especially valuable for tasks like populating databases, creating summaries, or generating reports from large amounts of text data.

Entity Categorization: NER categorizes entities into predefined classes such as persons, organizations, locations, dates, and more. This categorization provides context and semantic meaning, enabling more sophisticated analysis of the text.

Improved Search and Retrieval: By identifying and tagging named entities, NER can improve the accuracy and relevance of search results in applications like search engines, document retrieval systems, and recommendation systems.

Question Answering: NER is essential for question answering systems, as it helps identify relevant entities in the text that are related to the question being asked. This can lead to more accurate and informative answers.

Entity Linking: NER can be used to link recognized entities to knowledge bases, such as Wikipedia or other domain-specific databases, enriching the information by connecting it to external resources.

Named Entity Disambiguation: NER can help in disambiguating the context of an entity. For instance, the same name "Apple" could refer to a fruit or a technology company. NER can help distinguish between these different meanings based on the context.

Sentiment Analysis: In sentiment analysis, recognizing entities can help determine the sentiment towards specific entities. This is useful for understanding public opinion about companies, products, or individuals.

Event Extraction: NER can aid in extracting events and relationships between entities in a text. This is useful for tasks like event detection, timeline generation, and understanding connections between entities.

Regulatory Compliance and Data Security: In industries like finance and healthcare, NER can assist in identifying sensitive information, like personal names, medical terms, and financial figures, ensuring compliance with data protection regulations.

Language Translation and Generation: NER can improve the accuracy of machine translation and text generation by preserving the names of entities, resulting in more coherent and contextually relevant output.

Data Analysis and Visualization: By extracting and categorizing entities, NER can facilitate data analysis and visualization, making it easier to identify trends, patterns, and relationships in large text datasets.

Automated Document Summarization: NER can be used to identify key entities in a document, which in turn can be used to generate informative and concise document summaries.

Overall, NER is a foundational tool in NLP that adds structure, context, and meaning to text data, enabling a wide range of applications that require understanding and processing human language.

Code: -

Google Collaboratory Link: -

Exp 5-AI/ML

Conclusion: -

Comment on the role of Named Entity Recognition (NER) played in Natural Language Processing, and how it enhance the understanding and processing of unstructured text data.