Practical No:6

Title:

Implement Named Entity Recognizer.

Aim:

To implement a **Named Entity Recognizer (NER)** to identify and classify named entities in a given text, such as **person names**, **organizations**, **locations**, **dates**, etc.

Pre-requisites:

1. Basic understanding of **Named Entity Recognition (NER)**.
2. Familiarity with NLP libraries such as nltk or spaCy for NER.
3. Knowledge of machine learning models used in NER, like Conditional Random Fields (CRF) or pre-trained models for entity recognition.

Theory:

Named Entity Recognition (NER):

NER is the process of locating and classifying named entities in unstructured text into predefined categories, such as:

- **PERSON**: Names of people (e.g., "Albert Einstein").
- **ORGANIZATION**: Names of organizations (e.g., "Google").
- **LOCATION**: Geographic locations (e.g., "New York").
- **DATE/TIME**: Dates or times (e.g., "January 1st, 2024").
- **MONEY**: Monetary values (e.g., "$10 million").
- **PERCENTAGE**: Percentages (e.g., "30%").

NER is widely used in various NLP tasks such as information retrieval, question answering, and summarization.

How NER Works:

1. **Text Preprocessing:** Tokenize the text into sentences and words.
2. **POS Tagging:** Assign Part-of-Speech (POS) tags to words to provide syntactic information.
3. **Entity Recognition:** Use NER models to recognize named entities in the text and classify them into predefined categories.
4. **Contextual Understanding:** Some NER systems leverage pre-trained models that understand the context of a word in a sentence to better classify the entities.

Steps for Named Entity Recognition:

1. **Text Preprocessing:**
   ○ Tokenize the input text into sentences and words.
2. **POS Tagging (Optional):**
   ○ Use a POS tagger to assign grammatical categories to each word (e.g., noun, verb).
3. **Named Entity Recognition:**
   ○ Apply an NER model, such as one available in the nltk or spaCy libraries, to recognize named entities.
4. **Classify Named Entities:**
   ○ Entities are classified into categories like PERSON, ORGANIZATION, LOCATION, etc.

Conclusion:

Named Entity Recognition (NER) is a crucial task in natural language processing that enables systems to identify and classify important pieces of information from unstructured text. Implementing NER using libraries like `nltk` and `spaCy` simplifies the process as they provide pre-trained models for entity recognition. This technique is widely used in applications like search engines, recommendation systems, and information extraction.