

NLP Experiment 9

Aim: To implement Named Entity Recognition for a given real-world application.

Theory:

Named Entity Recognition (NER) is a crucial task in Natural Language Processing (NLP) that involves identifying and categorizing named entities into predefined classes such as persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc.

Implementing NER for a real-world application involves several steps, including data preparation, model selection, training, and evaluation.

Named Entity Recognition is done using Spacy's model. Entities that can be recognized with Spacy's model are:

Person: Names of individuals, including first and last names.

1. Organization: Names of companies, institutions, and organizations.
2. Location: Geographical places, such as cities, countries, states, and landmarks.
3. Date: Specific dates or date ranges, including days, months, and years.
4. Time: Time expressions, including specific times or time intervals.
5. Money: Monetary values, such as currency amounts and prices.
6. Percentage: Percentage values, such as percentages of change.
7. Cardinal Number: Numerical values, both cardinal (e.g., "one," "two") and numeric (e.g., "1," "2").
8. Ordinal Number: Ordinal numbers, indicating order or rank (e.g., "first," "second").
9. Quantity: Measurements and quantities, such as distances, weights, and volumes.
10. Language: Names of languages or language-specific terms.
11. Event: Names of specific events, conferences, or occurrences.
12. Law: Legal references, statutes, or legal terms.
13. Product: Names of products, goods, or branded items.
14. Work of Art: Titles of books, movies, songs, paintings, and other artistic works.
15. Drug: Names of pharmaceutical drugs and medications.
16. NORP (Nationalities or Religious/Political Groups): Names of nationalities, religious, or political groups.
17. Facility: Names of buildings, facilities, or physical structures.
18. Email Address: Email addresses or references to electronic mail.
19. Phone Number: Telephone numbers or references to phone communications.
20. URL: Web URLs or internet addresses.
21. GPE (Geopolitical Entity): Names of geopolitical entities, such as countries, cities, and regions.
22. Honorific (Title): Titles, honorifics, and forms of address (e.g., "Mr.," "Dr.").

- 23. Money Range: Ranges of monetary values.
- 24. Quantity Range: Ranges of quantities or measurements.
- 25. Time Range: Ranges of time expressions or intervals.

Libraries and Tools Used:

- Pandas
- Spacy