



Vidyavardhini's College of Engineering & Technology

Department of Computer Science and Engineering (Data Science)

---

Experiment No.7
Implement Named Entity Recognizer for the given Text input
Date of Performance:
Date of Submission:



# Vidyavardhini's College of Engineering & Technology

Department of Computer Science and Engineering (Data Science)

---

**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).



# Vidyavardhini's College of Engineering & Technology

## Department of Computer Science and Engineering (Data Science)

### Implementation:

```
Untitled.ipynb
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
Experiment 07 – NLP DLOC – Sahil Gujral – Roll No. 15 – Dept. of CSE (DS) – VCET

[1] In [1]: !pip install spacy
Python -> spacy download en_core_web_sm

Requirement already satisfied: spacy in /usr/local/lib/python3.10/dist-packages (3.6.1)
Requirement already satisfied: spacy-legacy<1.0.0,>=0.0.11 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.0.5)
Requirement already satisfied: murmurhash<1.0.0,>=0.28.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.0.10)
Requirement already satisfied: cytoolz<1.0.0,>=0.2.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.0.8)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.0.9)
Requirement already satisfied: thinc<8.2.0,>=8.1.8 in /usr/local/lib/python3.10/dist-packages (from spacy) (8.1.12)
Requirement already satisfied: wasabi<1.2.0,>=0.5.1 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.1.2)
Requirement already satisfied: srsly<1.0.0,>=2.4.3 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.4.8)
Requirement already satisfied: catalogue<1.0.0,>=0.4.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.0.10)
Requirement already satisfied: typer<0.10.0,>=0.3.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (0.9.0)
Requirement already satisfied: path<0.10.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (0.10.1)
Requirement already satisfied: smart-open<7.0.0,>=5.1.1 in /usr/local/lib/python3.10/dist-packages (from spacy) (6.4.0)
Requirement already satisfied: tqdm<0.0.0,>=4.38.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (4.66.1)
Requirement already satisfied: numpy<1.15.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.23.5)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.31.0)
Requirement already satisfied: pydantic<1.8.1,>=1.8.1,<3.0.0,>=1.7.4 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.10.13)
Requirement already satisfied: Jinja2 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.1.2)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from spacy) (67.7.2)
Requirement already satisfied: packaging<20.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (23.2)
Requirement already satisfied: langcodes<0.0,>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.3.0)
Requirement already satisfied: typing-extensions<4.2.0 in /usr/local/lib/python3.10/dist-packages (from pydantic<1.8.1,>=1.8.1,<3.0.0,>=1.7.4-spacy) (4.5.0)
Requirement already satisfied: charset-normalizer<4.0.0 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0-spacy) (3.3.1)
Requirement already satisfied: idna<4.0,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0-spacy) (3.4)
Requirement already satisfied: urllib3<3.0,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0-spacy) (2.0.7)
✓ 0s completed at 9:10PM
```

```
Untitled.ipynb
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from spacy) (67.7.2)
Requirement already satisfied: packaging<20.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (23.2)
Requirement already satisfied: langcodes<0.0,>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.3.0)
Requirement already satisfied: typing-extensions<4.2.0 in /usr/local/lib/python3.10/dist-packages (from pydantic<1.8.1,>=1.8.1,<3.0.0,>=1.7.4-spacy) (4.5.0)
Requirement already satisfied: charset-normalizer<4.0.0 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0-spacy) (3.3.1)
Requirement already satisfied: idna<4.0,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0-spacy) (3.4)
Requirement already satisfied: urllib3<3.0,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0-spacy) (2.0.7)
Requirement already satisfied: certifi<2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0-spacy) (2023.7.22)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in /usr/local/lib/python3.10/dist-packages (from thinc<8.2.0,>=8.1.8-spacy) (0.7.11)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.10/dist-packages (from thinc<8.2.0,>=8.1.8-spacy) (0.1.3)
Requirement already satisfied: click<9.0.0,>=7.1.1 in /usr/local/lib/python3.10/dist-packages (from typer<0.10.0,>=0.3.0-spacy) (8.1.7)
Requirement already satisfied: MarkupSafe<2.0 in /usr/local/lib/python3.10/dist-packages (from Jinja2-spacy) (2.1.3)
2023-10-11 15:38:14.362037: E tensorflow/compiler/xla/stream_executor/cuda/cuda_dnn.cc:1042] unable to register cudnn factory: Attempting to register factory for plugin cudnn when or
2023-10-11 15:38:14.362462: E tensorflow/compiler/xla/stream_executor/cuda/cuda_dnn.cc:1042] unable to register cudnn factory: Attempting to register factory for plugin cudnn when or
2023-10-11 15:38:14.372081: I tensorflow/core/platform/cpu_feature_guard.cc:182] This TensorFlow binary is optimized to use available CPU instructions in performance-critical operat
to enable the following instructions: AVX2 FMA, in other operations, rebuild TensorFlow with the appropriate compiler flags.
2023-10-11 15:38:15.528015: W tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: could not find TensorRT
Collecting en-core-web-sm==3.6.0
  Downloading https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-3.6.0/en_core_web_sm-3.6.0-py3-none-any.whl (12.8 MB)
    12.8/12.8 MB 88.9 MB/s eta 0:00:00
Requirement already satisfied: spacy<3.7.0,>=3.6.0 in /usr/local/lib/python3.10/dist-packages (from en-core-web-sm==3.6.0) (3.6.1)
Requirement already satisfied: spacy-legacy<1.0.0,>=0.0.11 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0-en-core-web-sm==3.6.0) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0-en-core-web-sm==3.6.0) (1.0.5)
Requirement already satisfied: murmurhash<1.0.0,>=0.28.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0-en-core-web-sm==3.6.0) (1.0.10)
Requirement already satisfied: cytoolz<1.0.0,>=0.2.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0-en-core-web-sm==3.6.0) (2.0.8)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0-en-core-web-sm==3.6.0) (3.0.9)
Requirement already satisfied: thinc<8.2.0,>=8.1.8 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0-en-core-web-sm==3.6.0) (8.1.12)
Requirement already satisfied: wasabi<1.2.0,>=0.5.1 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0-en-core-web-sm==3.6.0) (1.1.2)
Requirement already satisfied: srsly<1.0.0,>=2.4.3 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0-en-core-web-sm==3.6.0) (2.4.8)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0-en-core-web-sm==3.6.0) (2.0.10)
Requirement already satisfied: typer<0.10.0,>=0.3.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0-en-core-web-sm==3.6.0) (0.9.0)
Requirement already satisfied: path<0.10.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0-en-core-web-sm==3.6.0) (0.10.3)
✓ 0s completed at 9:10PM
```

```
Untitled.ipynb
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
[1] Requirement already satisfied: certifi<2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0-spacy<3.7.0,>=3.6.0-en-core-web-sm==3.6.0) (2023.7.22)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in /usr/local/lib/python3.10/dist-packages (from thinc<8.2.0,>=8.1.8-spacy<3.7.0,>=3.6.0-en-core-web-sm==3.6.0) (0.7.11)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.10/dist-packages (from thinc<8.2.0,>=8.1.8-spacy<3.7.0,>=3.6.0-en-core-web-sm==3.6.0) (0.1.3)
Requirement already satisfied: click<9.0.0,>=7.1.1 in /usr/local/lib/python3.10/dist-packages (from typer<0.10.0,>=0.3.0-spacy<3.7.0,>=3.6.0-en-core-web-sm==3.6.0) (8.1.7)
Requirement already satisfied: MarkupSafe<2.0 in /usr/local/lib/python3.10/dist-packages (from Jinja2-spacy<3.7.0,>=3.6.0-en-core-web-sm==3.6.0) (2.1.3)
✓ Download and installation successful
You can now load the package via spacy.load("en_core_web_sm")

[2] In [2]: import spacy

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

[4] In [4]: text = "Apple Inc. is planning to open a new store in New York City."

[5] In [5]: doc = nlp(text)

[6] In [6]: for entity in doc.ents:
      print(f'Entity: {entity.text}, Label: {entity.label}')

Entity: Apple Inc., Label: ORG
Entity: New York City, Label: GPE

from spacy.display.render import EntityRenderer
options = {"ents": ["ORG", "GPE"]}
entity_renderer(options=options).render(doc)
```

### Conclusion: