ANISH ADNANI
D17B/01

EXPERIMENT 8

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

Theory:

Named entity recognizer

Named Entity Recognition is a process where an algorithm takes a string of text as input and identifies relevant nouns (people, places, and organizations) that are mentioned in that string. It is sometimes referred to as entity chunking, extraction, or identification. It is task of identifying and categorizing key information in text.

For example, an NER machine learning model might detect the word "super.AI" in a text and classify it as a "company"

NER is a form of NLP, a subfield of artificial intelligence.

NER working

Any NER model consists of two steps

1. Detect a named entity
2. Categorize the entity

Step one involves detecting a word or string of words that form an entity. Each word represents a token "The Great Lakes" is a string of three tokens that represent one entity.

Inside-outside-beginning tagging is a common way of indicating where entities begin and end.

The second step involves the creation of entity categories: some common entity categories are person, organization, Time, location, work of art.

NER processes the structured and unstructured texts by identifying and locating the entities. It works like first identifying the entities and then also classifies them to allocate in a particular category like person, organization or location.

uses:

- Efficient search Algorithms:
  When designing a search algorithm for an online publisher that has millions of articles: If for every query algorithm looks at all million words then it will take a lot of time, instead NER can be run once on all the articles and the relevant entities (tags) associated with each of those articles are stored seperately.

- Powering content recommendations:
  This can be done by extracting entities from a particular article and recommending the other articles which have the most similar entities mentioned in them.

- Customer support:

There are a number of ways to make the process of customer feedback handling smooth and NER could be one of them. When handling the customer support department of an electronic store with multiple branches worldwide, we need to go through a series of names mentioned in customers feedback.

Conclusion:

We have successfully implemented Named Entity Recognizer for the given text input-text as for a normal sentence and for a medical organization

jupyter anish-nlp-exp8 Last Checkpoint: a minute ago (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run Code

In [1]:
```
import nltk
from nltk.tokenize import word_tokenize
from nltk.tag import pos_tag
from pprint import pprint
```

In [2]:
```
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')
nltk.download('maxent_ne_chunker')
nltk.download('words')
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]       date!
[nltk_data] Downloading package maxent_ne_chunker to
[nltk_data]     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping chunkers\maxent_ne_chunker.zip.
[nltk_data] Downloading package words to
[nltk_data]     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping corpora\words.zip.
```

Out[2]: True

In [3]:
```
sentence = "Researchers at the three universities in south korea developed an artifical intelligent model"
```

In [4]:
```
def preprocess(sent_tokens):
    sent_tokens = nltk.word_tokenize(sent_tokens)
    sent_tokens = nltk.pos_tag(sent_tokens)
    return sent_tokens
sent_tokens = preprocess(sentence)
sent_tokens
```

Out[4]:
```
[('Researchers', 'NNS'),
 ('at', 'IN'),
 ('the', 'DT'),
 ('three', 'CD'),
 ('universities', 'NNS'),
 ('in', 'IN'),
 ('south', 'JJ'),
 ('korea', 'NN'),
 ('developed', 'VBD'),
 ('an', 'DT'),
 ('artifical', 'JJ'),
 ('intelligent', 'NN'),
 ('model', 'NN')]
```

In [5]:
```
pattern = 'NP: {<DT>?<JJ>*<NN>}'
cp = nltk.RegexpParser(pattern)
cs = cp.parse(sent_tokens)
print(cs)
```

```
(S
  Researchers/NNS
  at/IN
  the/DT
  three/CD
  universities/NNS
  in/IN
  (NP south/JJ korea/NN)
  developed/VBD
  (NP an/DT artifical/JJ intelligent/NN)
  (NP model/NN))
```

In [6]:
```
ne_tree = nltk.ne_chunk(pos_tag(word_tokenize(sentence)))
print(ne_tree)
```

```
(S
  Researchers/NNS
  at/IN
  the/DT
  three/CD
  universities/NNS
  in/IN
  south/JJ
  korea/NN
  developed/VBD
  an/DT
  artifical/JJ
  intelligent/NN
  model/NN)
```

NER Using Spacy

In [13]:
```
#NER Using Spacy
import spacy
from spacy import displacy
from collections import Counter
import en_core_web_sm
nlp = en_core_web_sm.load()
from pprint import pprint
```

In [15]:
```
doc = nlp("Researchers at the three universities in south korea developed an artifical intelligent model")
pprint([(X.text, X.label ) for X in doc.ents])
```

```
[('three', 397)]
```

In [16]:
```
pprint([(X, X.ent_iob_, X.ent_type_) for X in doc])
```

```
[(Researchers, 'O', ''),
 (at, 'O', ''),
 (the, 'O', ''),
 (three, 'B', 'CARDINAL'),
 (universities, 'O', ''),
 (in, 'O', ''),
 (south, 'O', ''),
 (korea, 'O', ''),
 (developed, 'O', ''),
 (an, 'O', ''),
 (artifical, 'O', ''),
 (intelligent, 'O', ''),
 (model, 'O', '')]
```

In [19]:
```
entities= nlp(sentence)
```

In [22]:
```
print("Named entities in this text are:\n")
for ent in doc.ents:
    print(ent.text,ent.label_)
```

```
Named entities in this text are:

three CARDINAL
```

In [23]:
```
displacy.render(entities, style='ent', jupyter=True)
```

Researchers at the three CARDINAL universities in south korea developed an artifical intelligent model

NER Using Spacy for articles

In [24]:
```
import spacy
nlp=spacy.load('en_core_web_sm')
nlp.pipe_names
```

Out[24]:
```
['tok2vec', 'tagger', 'parser', 'ner', 'attribute_ruler', 'lemmatizer']
```

In [26]:
```
article_text="Researchers at three universities in South Korea have developed an artificial system stimulating a conscious respor
doc=nlp(article_text)
for ent in doc.ents:
    print(ent.text,ent.label_)
```

```
three CARDINAL
South Korea GPE
```

In [27]:
```
displacy.render(doc, style='ent', jupyter=True)
```

Researchers at three CARDINAL universities in South Korea GPE have developed an artificial system stimulating a conscious response to external stimuli.

NER for Medical Information using SpaCy

```
pip install spacy
```

```
Requirement already satisfied: spacy in /usr/local/lib/python3.7/dist-packages (2.2.4
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.7/di
Requirement already satisfied: thinc==7.4.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in /usr/local/lib/python3.7/dist-
Requirement already satisfied: blis<0.5.0,>=0.4.0 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: plac<1.2.0,>=0.9.6 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in /usr/local/lib/python3.7/di
Requirement already satisfied: numpy>=1.15.0 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.7/dist
Requirement already satisfied: srsly<1.1.0,>=1.0.2 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.7/
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr/local/
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: importlib-metadata>=0.20; python_version < "3.8" in /u
Requirement already satisfied: typing-extensions>=3.6.4; python_version < "3.8" in /u
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (f
```

```
medical_info = "Malaria is a life-threatening disease caused by parasites that are transmi
```

```
pip install -U spacy
```

```
Requirement already satisfied, skipping upgrade: jinja2 in /usr/local/lib/python3.
Requirement already satisfied, skipping upgrade: typing-extensions<4.0.0.0,>=3.7.4
Requirement already satisfied, skipping upgrade: setuptools in /usr/local/lib/pyth
Collecting pydantic<1.8.0,>=1.7.1
  Downloading https://files.pythonhosted.org/packages/b3/0a/52ae1c659fc08f13dd7c0a
     |████████████████████████████████| 9.1MB 47.7MB/s
Requirement already satisfied, skipping upgrade: importlib-metadata>=0.20; python_
Collecting pathy>=0.3.5
  Downloading https://files.pythonhosted.org/packages/a2/53/97dc0197cca9357369b3b7
Requirement already satisfied, skipping upgrade: cymem<2.1.0,>=2.0.2 in /usr/local
Requirement already satisfied, skipping upgrade: tqdm<5.0.0,>=4.38.0 in /usr/local
Collecting catalogue<2.1.0,>=2.0.1
  Downloading https://files.pythonhosted.org/packages/82/a5/b5021c74c04cac35a27d34
Collecting spacy-legacy<3.1.0,>=3.0.0
  Downloading https://files.pythonhosted.org/packages/78/d8/e25bc7f99877de34def57d
Requirement already satisfied, skipping upgrade: murmurhash<1.1.0,>=0.28.0 in /usr
Requirement already satisfied, skipping upgrade: numpy>=1.15.0 in /usr/local/lib/p
Requirement already satisfied, skipping upgrade: requests<3.0.0,>=2.13.0 in /usr/l
Requirement already satisfied, skipping upgrade: wasabi<1.1.0,>=0.8.1 in /usr/loca
Requirement already satisfied, skipping upgrade: preshed<3.1.0,>=3.0.2 in /usr/loc
Collecting typer<0.4.0,>=0.3.0
  Downloading https://files.pythonhosted.org/packages/90/34/d138832f6945432c638f32
Collecting thinc<8.1.0,>=8.0.2
  Downloading https://files.pythonhosted.org/packages/e3/08/20e707519bcded1a0caa6f
     |████████████████████████████████| 1.1MB 36.0MB/s
```

```
Requirement already satisfied, skipping upgrade: pyparsing>=2.0.2 in /usr/local/li
Requirement already satisfied, skipping upgrade: MarkupSafe>=0.23 in /usr/local/li
Requirement already satisfied, skipping upgrade: zipp>=0.5 in /usr/local/lib/pytho
Collecting smart-open<4.0.0,>=2.2.0
  Downloading https://files.pythonhosted.org/packages/11/9a/ba2d5f67f25e8d5bbf2fce
     |████████████████████████████████| 122kB 47.5MB/s
Requirement already satisfied, skipping upgrade: chardet<4,>=3.0.2 in /usr/local/l
Requirement already satisfied, skipping upgrade: idna<3,>=2.5 in /usr/local/lib/py
Requirement already satisfied, skipping upgrade: certifi>=2017.4.17 in /usr/local/
Requirement already satisfied, skipping upgrade: urllib3!=1.25.0,!=1.25.1,<1.26,>=
Requirement already satisfied, skipping upgrade: click<7.2.0,>=7.1.1 in /usr/local
Building wheels for collected packages: smart-open
  Building wheel for smart-open (setup.py) ... done
  Created wheel for smart-open: filename=smart_open-3.0.0-cp37-none-any.whl size=1
  Stored in directory: /root/.cache/pip/wheels/18/88/7c/f06dabd5e9cabe02d2269167bc
Successfully built smart-open
Installing collected packages: catalogue, srsly, pydantic, typer, smart-open, path
  Found existing installation: catalogue 1.0.0
    Uninstalling catalogue-1.0.0:
      Successfully uninstalled catalogue-1.0.0
  Found existing installation: srsly 1.0.5
    Uninstalling srsly-1.0.5:
      Successfully uninstalled srsly-1.0.5
  Found existing installation: smart-open 5.0.0
    Uninstalling smart-open-5.0.0:
      Successfully uninstalled smart-open-5.0.0
  Found existing installation: thinc 7.4.0
    Uninstalling thinc-7.4.0:
      Successfully uninstalled thinc-7.4.0
  Found existing installation: spacy 2.2.4
    Uninstalling spacy-2.2.4:
      Successfully uninstalled spacy-2.2.4
Successfully installed catalogue-2.0.3 pathy-0.4.0 pydantic-1.7.3 smart-open-3.0.0
```

```
pip install scispacy
```

```
Collecting scispacy
  Downloading https://files.pythonhosted.org/packages/46/d2/456e1f66f7ba65209746aac6
     |████████████████████████████████| 51kB 3.7MB/s
Requirement already satisfied: numpy in /usr/local/lib/python3.7/dist-packages (from
Collecting pysbd
  Downloading https://files.pythonhosted.org/packages/48/0a/c99fb7d7e176f8b176ef1970
     |████████████████████████████████| 71kB 4.3MB/s
Requirement already satisfied: scikit-learn>=0.20.3 in /usr/local/lib/python3.7/dist-
Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: spacy<3.1.0,>=3.0.0 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: requests<3.0.0,>=2.0.0 in /usr/local/lib/python3.7/dis
Collecting conllu
  Downloading https://files.pythonhosted.org/packages/ae/be/be6959c3ff2dbfdd87de4be0c
Collecting nmslib>=1.7.3.6
  Downloading https://files.pythonhosted.org/packages/be/77/aebbd03a32488024d2ae2230b
     |████████████████████████████████| 13.5MB 317kB/s
Requirement already satisfied: scipy>=0.17.0 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: thinc<8.1.0,>=8.0.2 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: pathy>=0.3.5 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: srsly<3.0.0,>=2.4.0 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: typer<0.4.0,>=0.3.0 in /usr/local/lib/python3.7/dist-p
```

```
Requirement already satisfied: catalogue<2.1.0,>=2.0.1 in /usr/local/lib/python3.7/di
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.7/dist
Requirement already satisfied: jinja2 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: typing-extensions<4.0.0.0,>=3.7.4; python_version < "3
Requirement already satisfied: importlib-metadata>=0.20; python_version < "3.8" in /u
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.0 in /usr/local/lib/python3.7
Requirement already satisfied: wasabi<1.1.0,>=0.8.1 in /usr/local/lib/python3.7/dist-
Requirement already satisfied: pydantic<1.8.0,>=1.7.1 in /usr/local/lib/python3.7/dis
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.7/
Requirement already satisfied: blis<0.8.0,>=0.4.0 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr/local/
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: psutil in /usr/local/lib/python3.7/dist-packages (from
Collecting pybind11<2.6.2
  Downloading https://files.pythonhosted.org/packages/00/84/fc9dc13ee536ba5e6b8fd10ce
     |████████████████████████████████| 194kB 44.5MB/s
Requirement already satisfied: pyparsing>=2.0.2 in /usr/local/lib/python3.7/dist-pack
Requirement already satisfied: smart-open<4.0.0,>=2.2.0 in /usr/local/lib/python3.7/d
Requirement already satisfied: click<7.2.0,>=7.1.1 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: zipp>=0.5; python_version < "3.8" in /usr/local/lib/py
Requirement already satisfied: MarkupSafe>=0.23 in /usr/local/lib/python3.7/dist-pack
Installing collected packages: pysbd, conllu, pybind11, nmslib, scispacy
Successfully installed conllu-4.4 nmslib-2.1.1 pybind11-2.6.1 pysbd-0.3.4 scispacy-0
```

```
pip install https://s3-us-west-2.amazonaws.com/ai2-s2-scispacy/releases/v0.4.0/en_ner_bc5c
```

```
Collecting https://s3-us-west-2.amazonaws.com/ai2-s2-scispacy/releases/v0.4.0/en_ner_
  Downloading https://s3-us-west-2.amazonaws.com/ai2-s2-scispacy/releases/v0.4.0/en_n
     |████████████████████████████████| 125.1MB 85kB/s
Requirement already satisfied: spacy<3.1.0,>=3.0.1 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.7/
Requirement already satisfied: typing-extensions<4.0.0.0,>=3.7.4; python_version < "3
Requirement already satisfied: srsly<3.0.0,>=2.4.0 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: typer<0.4.0,>=0.3.0 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: numpy>=1.15.0 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: wasabi<1.1.0,>=0.8.1 in /usr/local/lib/python3.7/dist-
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: jinja2 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: pydantic<1.8.0,>=1.7.1 in /usr/local/lib/python3.7/dis
Requirement already satisfied: thinc<8.1.0,>=8.0.2 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.7/di
Requirement already satisfied: blis<0.8.0,>=0.4.0 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: importlib-metadata>=0.20; python_version < "3.8" in /u
Requirement already satisfied: pathy>=0.3.5 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.7/dist
Requirement already satisfied: catalogue<2.1.0,>=2.0.1 in /usr/local/lib/python3.7/di
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.0 in /usr/local/lib/python3.7
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: click<7.2.0,>=7.1.1 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: pyparsing>=2.0.2 in /usr/local/lib/python3.7/dist-pack
Requirement already satisfied: MarkupSafe>=0.23 in /usr/local/lib/python3.7/dist-pack
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-pa
```

```
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr/local/
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (f
Requirement already satisfied: smart-open<4.0.0,>=2.2.0 in /usr/local/lib/python3.7/d
Building wheels for collected packages: en-ner-bc5cdr-md
  Building wheel for en-ner-bc5cdr-md (setup.py) ... done
  Created wheel for en-ner-bc5cdr-md: filename=en_ner_bc5cdr_md-0.4.0-cp37-none-any.w
  Stored in directory: /root/.cache/pip/wheels/fd/a0/5f/8e76c330a6c33638db0f177d6b596
Successfully built en-ner-bc5cdr-md
Installing collected packages: en-ner-bc5cdr-md
Successfully installed en-ner-bc5cdr-md-0.4.0
```

```
import scispacy
import spacy
import en_ner_bc5cdr_md   #The model we are going to use
from spacy import displacy
from scispacy.abbreviation import AbbreviationDetector
from scispacy.umls_linking import UmlsEntityLinker
```

```
nlp = spacy.load("en_ner_bc5cdr_md")
```

```
doc = nlp(medical_info)

print(list(doc.sents))
```

```
[Malaria is a life-threatening disease caused by parasites that are transmitted to pe
```

```
print(*doc.ents, sep = '\n')
```

```
Malaria
life-threatening disease
bites of infected female Anopheles mosquitoes
curable
malaria
malaria
malaria
malaria
malaria
malaria
deaths
```

```
from spacy import displacy
displacy.render((doc.sents), style='ent', jupyter=True)
```

Malaria DISEASE is a life-threatening disease DISEASE caused by parasites that are transmitted to people through the bites of infected female Anopheles mosquitoes DISEASE .

It is preventable and curable DISEASE .

```
doc = nlp(medical_info)

entity = doc.ents[3]
entity
```

```
curable
```