



Batch :- D-2 **Roll No. :-** 16010122151

Experiment :- 08

TITLE : To perform NLP on clinical data

AIM: Named Entity Recognition on Healthcare Data

Expected OUTCOME of Experiment:

CO5: Apply data analytics in the field of Health care.

Books/ Journals/ Websites referred:

Students have to list.

<https://www.cdc.gov/flu/treatment/antiviral-drugs.html>

Pre Lab/ Prior Concepts:

Students should have a basic understanding of natural language processing concepts like named entity recognition.

Procedure:

Data set Used: Pretrained spaCy model

Step1: Installed required library

(Students should write the code and output)

```
!pip install spacy
!pip install
https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-3.5.0/en_core_web_sm-3.5.0.tar.gz
```



K. J. Somaiya College of Engineering, Mumbai-77 (A Constituent College of Somaiya Vidyavihar University) Department of Computer Engineering

```

pip install spacy
pip install https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-3.5.0/en_core_web_sm-3.5.0.tar.gz

Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.6.0,>=3.5.0->en_core_web_sm=3.5.0) (4.66.5)
Requirement already satisfied: numpy<1.15.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.6.0,>=3.5.0->en_core_web_sm=3.5.0) (1.26.4)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.6.0,>=3.5.0->en_core_web_sm=3.5.0) (2.32.3)
Collecting pydantic<1.8.1,>=1.8.1,<1.11.0,>=1.7.4 (from spacy<3.6.0,>=3.5.0->en_core_web_sm=3.5.0)
Collecting pydantic-1.10.18-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (152 kB)
Collecting pathlib-abc<=0.1.1 (from pathy<=10.0.0->spacy<3.6.0,>=3.5.0->en_core_web_sm=3.5.0)
Requirement already satisfied: Jinja2 in /usr/local/lib/python3.10/dist-packages (from spacy<3.6.0,>=3.5.0->en_core_web_sm=3.5.0) (3.1.4)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from spacy<3.6.0,>=3.5.0->en_core_web_sm=3.5.0) (75.1.0)
Requirement already satisfied: packaging<20.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.6.0,>=3.5.0->en_core_web_sm=3.5.0) (24.1)
Requirement already satisfied: language-data<=1.2 in /usr/local/lib/python3.10/dist-packages (from spacy<3.6.0,>=3.5.0->en_core_web_sm=3.5.0) (1.2.0)
Requirement already satisfied: typing-extensions<=4.2.0 in /usr/local/lib/python3.10/dist-packages (from pydantic<1.8.1,>=1.8.1,<1.11.0,>=1.7.4->spacy<3.6.0,>=3.5.0->en_core_web_sm=3.5.0) (4.12.2)
Collecting pathlib-abc-0.1.1-py3-none-any.whl.metadata (18 kB)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.6.0,>=3.5.0->en_core_web_sm=3.5.0) (3.4.0)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.6.0,>=3.5.0->en_core_web_sm=3.5.0) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.6.0,>=3.5.0->en_core_web_sm=3.5.0) (2.2.3)
Requirement already satisfied: certifi<2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.6.0,>=3.5.0->en_core_web_sm=3.5.0) (2024.8.30)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in /usr/local/lib/python3.10/dist-packages (from thinc<8.2.0,>=8.1.0->spacy<3.6.0,>=3.5.0->en_core_web_sm=3.5.0) (0.7.11)
Requirement already satisfied: confection<1.0.0,>=0.8.1 in /usr/local/lib/python3.10/dist-packages (from thinc<8.2.0,>=8.1.0->spacy<3.6.0,>=3.5.0->en_core_web_sm=3.5.0) (0.1.5)
Requirement already satisfied: click<9.0.0,>=7.1.1 in /usr/local/lib/python3.10/dist-packages (from typer<0.10.0,>=0.8.2->spacy<3.6.0,>=3.5.0->en_core_web_sm=3.5.0) (8.1.7)
Requirement already satisfied: MarkupSafe<2.0 in /usr/local/lib/python3.10/dist-packages (from Jinja2->spacy<3.6.0,>=3.5.0->en_core_web_sm=3.5.0) (2.1.0)
Requirement already satisfied: marisa-trie<0.7.7 in /usr/local/lib/python3.10/dist-packages (from language-data<=1.2->language-data<=1.2->spacy<3.6.0,>=3.5.0->en_core_web_sm=3.5.0) (1.2.1)
Downloading spacy-3.5.4-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (6.6 MB)
6.6/6.6 MB 40.8 MB/s eta 0:00:00
Downloading pathy-0.11.0-py3-none-any.whl (47 kB)
47.3/47.3 kB 2.4 MB/s eta 0:00:00
Downloading pathlib-abc-0.1.1-py3-none-any.whl (21 kB)
21.0/21.0 kB 55.1 MB/s eta 0:00:00
Downloading pydantic-1.10.18-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (3.1 MB)
3.1/3.1 MB 55.1 MB/s eta 0:00:00
Downloading smart-open-6.4.0-py3-none-any.whl (57 kB)
57.0/57.0 kB 4.0 MB/s eta 0:00:00
Downloading thinc-8.1.12-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (919 kB)
919.6/919.6 kB 33.2 MB/s eta 0:00:00
Downloading typer-0.9.4-py3-none-any.whl (45 kB)
45.0/45.0 kB 40.8 MB/s eta 0:00:00

```

Step2: Load pretrained spaCy model (Students should write the code and output)

```

[2] import spacy
nlp = spacy.load("en_core_web_sm")

```

Step 3: Add Sample healthcare-related text (Students should write the code and output) <https://www.cdc.gov/flu/treatment/antiviral-drugs.html>

```

# Sample healthcare-related text
healthcare_text = """

Treating Flu with Antiviral Drugs
At a glance
Flu antiviral drugs are prescription medicines that can be used to treat flu illness.
They can lessen symptoms and shorten the time you are sick.
Antiviral drugs work best when started within 1 to 2 days after flu symptoms begin. Loading...
CDC recommends prompt treatment for people who have flu or suspected flu and who are at increased risk of serious flu complications, such as pregnant people, people with asthma and chronic lung disease, diabetes (including gestation
antiviral used to treat flu
Treatment overview
Flu antiviral drugs are prescription medicines (pills, liquid, an inhaled powder, or an intravenous solution) that fight against flu viruses in your body. Antiviral drugs are not sold over the counter. You can only get them if you h
Treatment of Flu with Flu antiviral medications works best when started within 1-2 days after flu symptoms begin. Flu antiviral drugs can lessen symptoms and shorten the time you are sick by about a day. Starting antiviral treatment
Recommended antiviral drugs for this flu season
There are four FDA-approved antiviral drugs recommended by CDC to treat flu this season.
oseltamivir phosphate (available as a generic version or under the trade name Tamiflu®),
zanamivir (trade name Relenza®),
peramivir (trade name Rapivue®), and
baloxavir marboxil (trade name Xofluza®).
Oseltamivir
Generic oseltamivir and Tamiflu® are available as a pill or liquid suspension and are FDA approved for early treatment of flu in people 14 days and older.
Zanamivir
Zanamivir is a powdered medication that is inhaled and approved for early treatment of flu in people 7 years and older. Note: Zanamivir (trade name Relenza®) is administered using an inhaler device and is not recommended for people :
Peramivir
Peramivir is given once intravenously by a health care provider and is approved for early treatment of flu in people 6 months and older.
Baloxavir
Baloxavir is a pill given as a single dose by mouth and is approved for early treatment of flu in children 5 years to younger than 12 years who do not have any chronic medical conditions, and for all people 12 years and older. Note:
Who should take antiviral drugs
It's very important that flu antiviral drugs are started as soon as possible to treat patients who are:
hospitalized with flu,
people who are very sick with flu but who do not need to be hospitalized, and

```

Step4: Process Text (Students should write the code and output)



```
print(f"{ent.text} - {ent.label}")

# Print explanations of entity labels
print("\nEntity Label Explanations:")
for ent in doc.ents:
    print(f"{ent.text}: {spacy.explain(ent.label)}")
```

Entities in the text:
Treating Flu with Antiviral Drugs - WORK_OF_ART
1 to 2 days - DATE
CDC - ORG
COVID-19 - ORG
COVID-19 - ORG
1-2 days - DATE
about a day - DATE
four - CARDINAL
FDA - ORG
CDC - ORG
this season - DATE
Tamiflu® - PRODUCT
Relenza® - PERSON
Rapivab - PERSON
Xofluza® - ORG
Tamiflu® - ORG
FDA - ORG
14 days - DATE
Zanamivir
Zanamivir - PERSON
7 years - DATE
Zanamivir - PERSON
Relenza® - PERSON
five days - DATE
6 months - DATE
Baloxavir
Baloxavir - PERSON
5 years - DATE
12 years - DATE
12 years - DATE
Baloxavir - PERSON
Xofluza® - ORG

Entities in the text:
Treating Flu with Antiviral Drugs - WORK OF ART
1 to 2 days - DATE
CDC - ORG
COVID-19 - ORG
COVID-19 - ORG
1-2 days - DATE
about a day - DATE
four - CARDINAL
FDA - ORG
CDC - ORG
this season - DATE
Tamiflu® - PRODUCT
Relenza® - PERSON
Rapivab - PERSON
Xofluza® - ORG
Tamiflu® - ORG
FDA - ORG
14 days - DATE
Zanamivir
Zanamivir - PERSON
7 years - DATE
Zanamivir - PERSON
Relenza® - PERSON
five days - DATE
6 months - DATE
Baloxavir
Baloxavir - PERSON
5 years - DATE
12 years - DATE
12 years - DATE
Baloxavir - PERSON
Xofluza® - ORG



Baloxavir - PERSON
two days - DATE
five days - DATE
Learn - PRODUCT
CDC - ORG
the American Academy of Pediatrics - ORG
2 weeks old - DATE
Zanamivir - PERSON
7 years - DATE
6 months - DATE
Baloxavir - PERSON
5 years - DATE
Opening and Mixing Oseltamivir Capsules with Liquids if Child
Cannot Swallow Capsules
Pregnant - WORK_OF_ART
Baloxavir - PERSON
Antiviral - ORG
two days - DATE
Antibiotics - ORG
C. - NORP
Healthy Habits: - PERSON
Zanamivir - PERSON
the Food and Drug Administration - ORG
FDA - ORG
season-to-season - DATE
6 months - DATE
second - ORDINAL

Entity Label Explanations:
Treating Flu with Antiviral Drugs: Titles of books, songs, etc.
1 to 2 days: Absolute or relative dates or periods
CDC: Companies, agencies, institutions, etc.
COVID-19: Companies, agencies, institutions, etc.
COVID-19: Companies, agencies, institutions, etc.
1-2 days: Absolute or relative dates or periods
about a day: Absolute or relative dates or periods
four: Numerals that do not fall under another type
FDA: Companies, agencies, institutions, etc.
CDC: Companies, agencies, institutions, etc.
this season: Absolute or relative dates or periods
Tamiflu®: Objects, vehicles, foods, etc. (not services)
Relenza®: People, including fictional
Rapid: People, including fictional
Xofluza®: Companies, agencies, institutions, etc.
Tamiflu®: Companies, agencies, institutions, etc.
FDA: Companies, agencies, institutions, etc.
14 days: Absolute or relative dates or periods
Zanamivir
Zanamivir: People, including fictional
7 years: Absolute or relative dates or periods
Zanamivir: People, including fictional
Relenza®: People, including fictional



five days: Absolute or relative dates or periods
6 months: Absolute or relative dates or periods
Baloxavir: People, including fictional
Baloxavir: People, including fictional
5 years: Absolute or relative dates or periods
12 years: Absolute or relative dates or periods
12 years: Absolute or relative dates or periods
Baloxavir: People, including fictional
Xofluza®: Companies, agencies, institutions, etc.
Baloxavir: People, including fictional
two days: Absolute or relative dates or periods
five days: Absolute or relative dates or periods
Learn: Objects, vehicles, foods, etc. (not services)
CDC: Companies, agencies, institutions, etc.
the American Academy of Pediatrics: Companies, agencies, institutions, etc.
2 weeks old: Absolute or relative dates or periods
Zanamivir: People, including fictional
7 years: Absolute or relative dates or periods
6 months: Absolute or relative dates or periods
Baloxavir: People, including fictional
5 years: Absolute or relative dates or periods
Opening and Mixing Oseltamivir Capsules with Liquids if Child Cannot Swallow Capsules
Pregnant: Titles of books, songs, etc.
Baloxavir: People, including fictional
Antiviral: Companies, agencies, institutions, etc.
two days: Absolute or relative dates or periods
Antibiotics: Companies, agencies, institutions, etc.
C.: Nationalities or religious or political groups
Healthy Habits:: People, including fictional
Zanamivir: People, including fictional
the Food and Drug Administration: Companies, agencies, institutions, etc.
FDA: Companies, agencies, institutions, etc.
season-to-season: Absolute or relative dates or periods
6 months: Absolute or relative dates or periods
second: "first", "second", etc.

Step 5: Print named entities

(Students should write the code and output)



```
condition_labels = [
    "pregnant people", "hospitalized", "serious flu complications",
    "antiviral drugs", "antibiotics"
]

# Iterate through the text and identify entities based on the custom labels
entities = []

for label in disease_labels:
    if label.lower() in healthcare_text.lower():
        entities.append((label, "DISEASE"))

for label in drug_labels:
    if label.lower() in healthcare_text.lower():
        entities.append((label, "DRUG"))

for label in condition_labels:
    if label.lower() in healthcare_text.lower():
        entities.append((label, "CONDITION"))

# Print the identified entities
print("Entities in the text:")
for entity, label in entities:
    print(f"{entity} - {label}")
```

Entities in the text:
flu - DISEASE
asthma - DISEASE
chronic lung disease - DISEASE
diabetes - DISEASE
gestational diabetes - DISEASE
heart disease - DISEASE
COVID-19 - DISEASE
C. diff infection - DISEASE
oseltamivir - DRUG
Tamiflu - DRUG
zanamivir - DRUG
Relenza - DRUG
peramivir - DRUG
Rapivab - DRUG
baloxavir - DRUG
Xofluza - DRUG
pregnant people - CONDITION
hospitalized - CONDITION
serious flu complications - CONDITION
antiviral drugs - CONDITION
antibiotics - CONDITION

```
Entities in the text:
flu - DISEASE
asthma - DISEASE
chronic lung disease - DISEASE
diabetes - DISEASE
gestational diabetes - DISEASE
heart disease - DISEASE
COVID-19 - DISEASE
C. diff infection - DISEASE
oseltamivir - DRUG
Tamiflu - DRUG
zanamivir - DRUG
Relenza - DRUG
peramivir - DRUG
Rapivab - DRUG
baloxavir - DRUG
Xofluza - DRUG
pregnant people - CONDITION
hospitalized - CONDITION
serious flu complications - CONDITION
antiviral drugs - CONDITION
```



antibiotics - CONDITION

Students have to perform all the tasks illustrated above by adding other sample health care related text.

Date: _____

Signature of faculty in-charge

Post Lab Descriptive Questions:

Q.1 What is Natural Language Processing (NLP) and how is it applied to healthcare data?

Natural Language Processing (NLP) is a subfield of artificial intelligence (AI) focused on the interaction between computers and human language. NLP enables machines to understand, interpret, and generate human language in a valuable way. In the healthcare domain, NLP is particularly useful for extracting meaningful information from unstructured data, such as clinical notes, electronic health records (EHRs), research papers, and patient feedback.

Applications in healthcare:

- **Medical records analysis:** NLP can process clinical notes, extract important medical concepts (such as symptoms, diseases, and treatments), and structure them for further analysis.
- **Drug discovery:** NLP helps analyze large volumes of research papers and clinical trial data, speeding up drug discovery processes.
- **Patient management:** NLP assists in processing patient feedback and communication to improve care by detecting patient needs and concerns.
- **Public health:** NLP can monitor online discussions or social media for public health trends, enabling early detection of disease outbreaks or public health crises.

Q.2 What are named entities in the context of healthcare NLP, and how can Named Entity Recognition (NER) be useful in medical texts?



Named entities in healthcare NLP refer to specific terms related to medicine, such as diseases, drugs, symptoms, medical conditions, treatments, procedures, and more. For example, terms like "flu," "diabetes," "oseltamivir," and "heart disease" are named entities in medical texts.

Named Entity Recognition (NER) is an NLP technique used to automatically identify and classify these entities in text. In the context of healthcare, NER can be useful in the following ways:

- **Extracting key medical information:** NER can automatically pull out essential medical entities from clinical notes or medical records, such as patient diagnoses, medications, and treatment plans.
- **Medical research and literature:** NER can be applied to biomedical research papers to identify genes, proteins, diseases, and drugs, aiding in research analysis.
- **Enhancing searchability:** By tagging important entities in medical texts, NER improves the accuracy of search engines and knowledge retrieval in large medical databases or EHRs.

Q.3 What are some common applications of NLP in healthcare, such as clinical decision support, medical coding, or patient sentiment analysis?

NLP has a variety of applications in healthcare, each contributing to improved patient care, operational efficiency, and research. Here are some common use cases:

1. **Clinical decision support (CDS):**
 - o NLP processes clinical notes and patient records to provide doctors with relevant medical insights, such as potential diagnoses or treatment recommendations, based on patient symptoms and history. It helps in identifying drug interactions, monitoring disease progression, and flagging critical cases.
2. **Medical coding:**
 - o NLP helps automate the assignment of standardized medical codes (e.g., ICD codes) from clinical documentation. By extracting key terms such as diagnoses, procedures, and treatments, NLP systems can assign accurate codes, improving billing accuracy and reducing administrative workload.
3. **Patient sentiment analysis:**
 - o NLP is applied to analyze patient feedback, such as surveys, online reviews, or patient communication. By identifying sentiments, emotions, and concerns expressed by patients, healthcare providers can improve care quality and address patient dissatisfaction.



SOMAIYA
VIDYAVIHAR UNIVERSITY

K J Somaiya College of Engineering

K. J. Somaiya College of Engineering, Mumbai-77
(A Constituent College of Somaiya Vidyavihar University)
Department of Computer Engineering